NTP
National Toxicology Program
U.S. Department of Health and Human Services

NTP RESEARCH REPORT ON

NATIONAL TOXICOLOGY PROGRAM APPROACH TO GENOMIC DOSE-RESPONSE MODELING

NTP RR 5

APRIL 2018

NTP Research Report on National Toxicology Program Approach to Genomic Dose-Response Modeling

Research Report 5

National Toxicology Program

April 2018

Table of Contents

About This Report

National Toxicology Program[1]

[1]Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Collaborators

Scott Auerbach, Frederick Parham, Mary Wolfe, John Bucher

Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Scientific and technical expertise to development of the NTP approach and modeling software, ensuring inclusion of appropriate mathematical methods
Scott Auerbach, Ph.D.
Frederick Parham, Ph.D.

Oversight and implementation of activities related to the webinar series, external peer review, readability, and finalization of the NTP approach.
Mary Wolfe, Ph.D.

Scientific consultation, review, and support toward development of NTP's approach to genomic dose-response modeling
John Bucher, Ph.D.

Contributors

Division of Intramural Research, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Guidance on statistical and mathematical modeling methods
Pierre Bushel, Ph.D.
Keith Shockley, Ph.D.

National Center for Computational Toxicology, Environmental Protection Agency, Research Triangle Park, North Carolina, USA
Conceptual guidance methodology
Russell Thomas, Ph.D.

University of Pittsburgh, Graduate School of Public Health, Pittsburgh, Pennsylvania, USA
Guidance on statistical and mathematical modeling methods
Shyamal Pedadda, Ph.D. (formerly at NIEHS/DNTP)

Peer Review

NTP convened an expert panel on October 23–25, 2017, to peer review the *NTP Research Report on National Toxicology Program Approach to Genomic Dose-Response Modeling*. The panel members are listed below. These reviewers served as independent scientists, not as representatives of any institution, company, or governmental agency. In this capacity, the panel provided scientific input on NTP's proposed approach at the public meeting.

Prior to the expert panel meeting, NTP hosted four webinars that presented an overview of current approaches and best practices to genomic dose-response modeling and NTP's proposed response. The expert panel considered the information presented in the webinars, as well as other technical factors, during its peer review of NTP's proposed approach.

Meeting materials, the peer review report, and webinar recordings can be found on the NTP website: https://ntp.niehs.nih.gov/about/org/sep/ntpexpertpanel/past/index.html.

Peer Reviewers

Carole Yauk, Ph.D. (Chair)
Research Scientist, Environmental Health Science and Research Bureau
Health Canada
Ottawa, ON, Canada

Lyle Burgoon, Ph.D.
Leader, Bioinformatics and Computational Toxicology Group
U.S. Army Engineer Research and Development Center
Raleigh, NC

Rebecca Clewell, Ph.D.
Chief Scientific Officer
ScitoVation
Durham, NC

Ruili Huang, Ph.D.
Lead Informatics Scientist
Division of Pre-Clinical Innovation
National Center for Advancing Translational Sciences (NCATS)
National Institutes of Health (NIH)
Rockville, MD

Kamin Johnson, Ph.D.
Associate Toxicology Manager
The Dow Chemical Company
Midland, MI

Jorge Naciff, Ph.D.
Principal Scientist
Global Product Stewardship
Central Product Safety Division, Mason Business Center
Procter & Gamble
Mason, OH

Setia Pramana, Ph.D.
Associate Professor, Institute of Statistics
Adjunct Faculty, Department of Medical Epidemiology and Biostatistics
Karolinska Institutet
Jakarta, Indonesia

James Stevens, Ph.D.
Distinguished Research Fellow
Eli Lily and Company
Indianapolis, IN

Fred Wright, Ph.D.
Professor, Director of the Bioinformatics Research Center
North Carolina State University
Raleigh, NC

Publication Details

Abstract

A major component of the National Toxicology Program's effort to evolve the practice of toxicology from an observational science to one that is more mechanistic involves using functional omic technologies, such as transcriptomics. NTP's approach to genomic dose-response modeling focuses on identifying biological potency of the test article. The approach reflects recommendations of experts in toxicology and toxicogenomics. The steps in the process include 1) developing a study design that includes a broad range of doses that sufficiently describe the shape of the dose-response and maximizes the accuracy of benchmark dose (BMD) values derived for all responsive genes; 2) incorporating a signal detection filter to ensure the modeled data contain adequate signal to lend confidence that any effect is treatment related; 3) conducting effect size and trend tests to identify those genes exhibiting a biologically plausible and reproducible response to test article treatment; 4) fitting of parametric dose-response models derived from the EPA BMD software to identify a biological potency estimate (i.e., a BMD) for each gene exhibiting a dose-related response to treatment; 5) grouping of genes into predefined Gene Ontologies (GO), identifying the GO gene sets responding to treatment, and then determining the composite potency of the gene set (i.e., a gene-set BMD); and 6) providing a biologically and mechanistically focused interpretation. NTP has developed a software package, BMDExpress 2.0, which enables users to implement the approach and perform extensive data visualization. The approach to genomic dose-response modeling is anticipated to lead to greater consistency in reporting of genomic dose-response data and to facilitate the use of genomic dose-response data in risk assessment.

Introduction

Changes in the approach to toxicological assessment[1] and the advent of inexpensive, high-throughput transcriptomics data generation platforms have led to significant interest in the integration of genomic dose-response studies (GDRS) for determining test article biological potency. The focus on biological potency is related to observation that gene set-level potencies derived from GDRS generally provide a good approximation of toxicological potency (i.e., the most sensitive effects that are typically used as a point of departure for risk assessment) that have been determined using long-term, resource-intensive toxicity studies (e.g., carcinogenicity) in the same model organism (e.g., rat)[2]. Such an approximation of biological potency from short-term GDRS could significantly decrease the time required to estimate safe exposure levels and, when combined with estimated human exposure levels, can be used for prioritization based on margin of exposure for more in-depth testing.

The National Toxicology Program (NTP) proposes the use of the approach outlined here for both in vivo and in vitro genomic dose-response studies. Although the case for using such an approach in the context of in vivo studies seems straightforward and has been demonstrated in several publications[3-13], how in vitro studies will be used in a similar context is less clear. A critical additional component of the analysis of in vitro GDRS data required to translate the results into the current context of a regulatory setting will be in vitro-to-in vivo extrapolation (IVIVe). IVIVe uses bioactive in vitro concentrations to estimate external doses in an intact organism that would be predicted to produce a biological effect. Discussion of IVIVe technology is beyond the scope of this document but has been discussed extensively in recent publications[14]. Overall, NTP is aware of the additional challenges associated with in vitro GDRS within the context of influencing regulatory decision making and will make a concerted effort to address these issues in future research.

Independent of the challenges associated with in vivo or in vitro GDRS, many questions exist regarding how best to design, perform, and interpret these studies and to facilitate integration of these types of assessments into the safety assessment paradigm. Consensus or, at a minimum, guidance on how to carry out and analyze such studies would be helpful in advancing the utility of the findings in decision making regarding safety and exposure limits. This document describes a framework for performing genomic dose-response analysis that is largely consistent with the published approach to analysis embodied in the BMDExpress software[3; 15]. The described approach aims to develop screening-level assessments of test articles for use in prioritizing and setting interim exposure limits.

NTP convened an expert panel on October 23–25, 2017, at the National Institute of Environmental Health Sciences, Research Triangle Park, NC to obtain input on specific details of its proposed approach to genomic dose-response modeling. NTP carefully considered the panel's input and recommendations in determining what changes to the approach might be needed prior to finalization and has herein modified the proposed approach accordingly. NTP also will continue to monitor the scientific literature for the development of improved approaches to data modeling and analysis. Importantly, in considering the proposed approach, readers are asked to keep in mind that NTP's goal in pursuing GDRS is *not* to perform toxicogenomic-based hazard identification but instead to develop a testing tool that can quickly and cost effectively provide an approximation of biological potency that tracks closely with toxicological potencies used as points of departure in the traditional risk assessment paradigm.

1

We envision the NTP-proposed approach as a necessary step in considering the use of GDRS-based biological potency estimates for future use in more traditional risk assessment.

Overview of NTP's Approach

Approach: NTP's approach is generally consistent with an approach to genomic dose-response modeling outlined by Thomas et al.[3] and used extensively by other researchers with various slight modifications[4-13]. This approach entails (1) initial determination of an adequate signal in the data for modeling; (2) filtering the measured features (genes/probe sets) to remove those not responding to chemical treatment; (3) fitting each filtered feature to multiple, parametric, dose-response models, identifying the best-fit model, and deriving a potency value (BMD) from that model for the feature[16]; (4) parsing the features into predefined gene sets (e.g., Gene Ontology Biological Processes[17]); and (5) determining potency for each adequately populated gene set by deriving the median potency of the genes in each set. BMDExpress is a software package that facilitates data analysis in accordance with the outlined approach[15]. NTP recently modified the software to produce the new, freely available version, BMDExpress 2.0.[a]

Two additional issues, which are not immediately central to the data modeling pipeline but are critical to overall success of the genomic dose-response approach, are study design and biological interpretation of findings. With this in mind, NTP will employ the study-design approach of Slob et al.[18], in which biological samples are distributed over a broad dose range, allowing for more accurate estimates of model parameters. This design contrasts with a more traditional toxicology study design that includes a limited number of dose levels combined with a high level of biological replication. The addition of other study-design parameters (detailed below), such as study duration, organ/tissue/cell line selection, and others, will be based on published findings and consistent with guideline studies, where appropriate. To facilitate interpretation of results from genomic dose-response analysis, NTP will employ a variety of gene sets curated in the MsigDB Database that describe or embody a wide variety of signaling and biological processes. Initially, NTP plans to limit the interpretation of the findings to information curated by external sources with the goal of avoiding inaccurate associations with traditional toxicological hazard that could be misconstrued. Finally, for quantitative interpretation, NTP will use the genomic dose-response results from the most sensitive gene set based on median BMD when identifying the lowest dose at which biological changes occur (i.e., biological effect point of departure [BEPOD]). Based on previous work, NTP anticipates the BEPOD will provide a reasonable approximation (within a half log) of the most sensitive toxicological effect level from longer-term studies in the same model (e.g., rat or mouse).

To illustrate the envisioned use of genomic dose-response studies at NTP, we present the results of the BEPOD identification for multiple, 5-day, repeat-dose studies in rats (Appendix A). To allow for comparison with results from the GDRS, the most sensitive neoplastic and nonneoplastic endpoints from either the 2-year studies or subchronic studies were modeled to determine BMDs.

Justification for the approach: The primary justifications for the overall approach are historical precedent and peer review, as several research groups in government, academic, and private sectors have used the general approach, which has been peer reviewed[3-13]. Further, this approach to dose-response modeling is largely consistent with standards the U.S. Environmental Protection Agency (US EPA) uses to evaluate data from guideline toxicological assessments and to make regulatory decisions. US EPA methods have been extensively documented and

---

[a] https://github.com/auerbachs/BMDExpress-2.0/releases.

reviewed[b]. The design of the in-life portion of the proposal is focused on achieving the most accurate estimation of biological potencies within the context of the BMD analysis paradigm for all features that change with treatment[18]. NTP's mission includes providing data for regulatory decision making; thus, we hope that employing methods already accepted by much of the regulatory community will remove one major barrier to using genomic data in a regulatory context. Finally, results derived from GDRS of multiple chemicals, which used this approach, yielded estimated potency values that are similar to potency estimates derived from apical toxicological endpoints[19]. This finding is critical to the approach, as it relates to the translation of findings from GDRS for determining safety-focused interim exposure limits and for evaluating how the proposed approach can be used to support more standard risk assessments in the future[2].

NTP's approach to study design focuses on obtaining the best data to determine accurate estimates of biological potency using modeling. The use of a broad array of gene sets such as those curated in MSigDB is to ensure that all known biological and signaling processes are covered, therefore ensuring the most sensitive estimation of biological potency. Finally, reporting of the most sensitive gene-set BMD as the BEPOD is based on observations that the BMD for the most sensitive gene set from short-term, in vivo, GDRS closely approximates the most sensitive BMD values from guideline toxicological assessments of the same test article.

Despite the relatively broad use of the methods in the NTP analysis pipeline, questions remain about specific aspects of the analysis performed inconsistently across the published literature. The following section presents a detailed description of the steps in the analysis and the rationale behind the selection of specific settings and parameters.

---

[b]https://www.epa.gov/sites/production/files/2015-01/documents/benchmark_dose_guidance.pdf.

A Detailed Description of the Steps in the Genomic Dose-Response Analysis

Determining Adequate Signal in the Data

Approach: In the first step of the genomic dose-response analysis protocol, we will apply a commonly used statistical analysis (ANOVA, with a Benjamini and Hochberg False Discovery Rate [FDR] <0.05) to determine globally whether the signal in the data is adequate to model and likely to yield minimally reproducible findings. At least one feature must pass the test to deem the data set worthy of further analysis.

Justification for approach: Use of a premodeling test to determine signal is essential to avoid modeling data with no statistically plausible signal. Modeling data with no statistically plausible signal is likely to yield unreproducible results with highly inaccurate estimates of BMD values. Use of FDR-adjusted ANOVA takes into account the variance intrinsic to the data set under consideration and is therefore an effective means of identifying data sets populated solely with noise. Importantly, efforts by the MicroArray Quality Control project indicate that such a filter is not ideal for maximizing reproducibility across laboratories. Thus, a distinct filter process is used in the next step in the analysis with the focus on achieving maximum reproducibility of genomic dose-response results.

Filtering of Measured Features

Approach: In the second step of the genomic dose-response analysis protocol, we will apply a statistical trend test (Williams' Trend Test[20]) and effect size (i.e., fold change relative to control threshold to be determined empirically for each technology) filter to each data set (one data set per chemical) to remove measured features (i.e., genes/transcripts/probe sets) from subsequent analysis that do not demonstrate a response to test article treatment. Due to technical differences in signal detection and the effect of different normalization processes, thresholds for the statistical and effect-size filter will be determined empirically for each transcriptomic platform (e.g., RNA-Seq, TempO-Seq[21], Affymetrix microarray) using an optimal combination of three metrics, including noise elimination (elimination of false signal), permissiveness (detection of true signal), and reproducibility (maximizing overlap in findings from independent study replication). Specifically upon embarking on the use of the new platform, an initial study will be performed three times independently. Each experiment will use a large number of vehicle control samples and a minimum of three test article treatments. To identify a noise elimination threshold, multiple permutations of statistical thresholds and effect sizes will be employed to perform a complete modeling (i.e., run through the analysis pipeline described here) of the "null signal" data sets derived from randomly sampled vehicle controls. Parameter thresholds that yield "active" (see below for definition) gene sets will be eliminated. Remaining parameter settings will be ranked by their degree of permissivity by quantifying the total differential effected features from experiments demonstrating a dose-related response to treatment. Parameter settings yielding the higher number of differentially expressed features will be given higher ranks than those yielding fewer. Independent reproducibility will be assessed to evaluate the percentage of overlapping "active" gene sets from each replicate experiment with the parameter settings that yield the highest overlap receiving the highest rank. The ranks provided by the noise elimination/

5

permissiveness and the reproducibility assessments will be combined to identify an optimal threshold for processing the data.

Justification for approach: The use of a trend test such as the Williams' Test is consistent with practices the US EPA recommends when performing BMD modeling.[c] Regarding genomic dose-response modeling, NTP acknowledges limitations to using a more traditional trend test such as the Williams' Test that identifies only monotonic trends. Due to the complexity embodied in genomic dose-response, biologically meaningful, nonmonotonic gene and gene set-level responses to treatment are possible. To address this concern, NTP proposes to evaluate the use of nonparametric tests such as those embodied in the Origene analysis tool. The goal in this evaluation will be to gauge biological plausibility of additional features identified by the nonparametric trend test.

The basis for combining a statistical and effect size filter is rooted in the observation from MicroArray Quality Control studies that found that a nominal p-value combined with the fold-change filter yields the highest cross-laboratory and cross-platform reproducibility when evaluating transcriptomic data[22]. Because simply a fold change filter can yield greater cross-lab reproducibility in certain cases, NTP will also evaluate the use of an effect-size filter in isolation using the three criteria described above.

Fitting Features to Dose-Response Models

Approach: In the third step of the genomic dose-response analysis, we will fit dose-response curves to each measured feature that exhibits a response to chemical treatment as determined by the filtering approach described above. Dose-response modeling will be performed as described previously, with minor modifications[3; 15]. To model the data, polynomial 2°, linear, power, Hill, and exponential 2, 3, 4, and 5 dose-response models will be fitted to the measured features. Equations describing each model are given below. Dose values used in the modeling will be on the linear scale and thus will include true "0" values. NTP's BMDExpress 2.0 software for performing dose-response modeling includes the US EPA model executables used in US EPA's BMDS software and will incorporate the constrained high-degree polynomial models and nonparametric modeling approaches in the future.[d] The specific details on each model are described in the BMDS software guidance document.[e]

Model equations

Equations for the models are taken from the BMDS software guidance document.[f] In all equations, μ is the mean response predicted by the model.

Polynomial model

$$\mu(dose) = \beta_0 + \beta_1\ dose + \beta_2\ dose^2 + \cdots + \beta_n\ dose^n$$

where *n* is the degree of the polynomial.

[c]https://www.epa.gov/sites/production/files/2015-01/documents/benchmark_dose_guidance.pdf.
[d]https://www.epa.gov/bmds/download-benchmark-dose-software-bmds-model-executables and https://www.epa.gov/bmds.
[e]https://www.epa.gov/sites/production/files/2015-01/documents/benchmark_dose_guidance.pdf.
[f]https://www.epa.gov/sites/production/files/2015-01/documents/benchmark_dose_guidance.pdf.

Linear model

The linear model is a special case of the polynomial model with n fixed at 1.

Power model

$$\mu(dose) = \gamma + \beta\ dose^{\delta}$$

where $0 < \gamma < 1$, $\beta \geq 0$, and $18 \geq \delta > 0$.

Hill model

$$\mu(dose) = \gamma + \frac{v\ dose^n}{k^n + dose^n}$$

Exponential model[a]

The four exponential models (2 through 5) are:

$$\mu(dose) = a * \exp(sign * b * dose)$$

$$\mu(dose) = a * \exp(sign * (b * dose)^d)$$

$$\mu(dose) = a * (c - (c - 1) * \exp(-1 * b * dose))$$

$$\mu(dose) = a * (c - (c - 1) * \exp(-1 * (b * dose)^d))$$

[a]*For the first two exponential models, 'sign' is the adverse direction.*

NTP will include additional higher degree poly models in the future when constraining the US EPA model executables becomes possible to allow change direction only once. Incorporation of these additional models will allow for more accurate modeling of a variety of nonmonotonic responses. NTP also intends to evaluate the use of nonparametric modeling approaches to determine if such approaches can be used to describe the shape of the diversity of dose-response behaviors observed in genomic dose-response with more accuracy. Nonparametric modeling approaches will be selected in consultation with the US EPA.

The data will be log2 adjusted and presumed to have a normal distribution; each model will be run assuming constant variance. In the future, a test for dose-related variance will be implemented for each feature, and nonconstant variance will be used in the modeling when appropriate.

The potency value derived from each model is the BMD[16].

- BMD is defined as the estimated dose or concentration that produces a predetermined change in the response rate of a biological response (called the benchmark response or BMR) compared to background[16].

- The BMR for each feature will be the standard deviation (SD) at zero dose. When constant variance is assumed, SD is estimated using the entire fitted curve, not just the control data. The use of 1 SD as a BMR is consistent with the proposed approach to modeling continuous data by the US EPA when the known, predefined adversity threshold is unknown. Currently, the degree of change associated with adversity in the gene expression space is largely unknown, thus the choice of 1 SD is appropriate. Note:

In the future, when the nonconstant variance assumption is selected and implemented for certain features, a BMR of 1 SD will be used and based solely on the variance at control, not the entire fitted curve as is done with the assumption of constant variance.

- An adverse direction (i.e., the direction of response indicating up- or down-regulation) is not selected a priori, but determined by a trend test embedded in the model executable. The selection of an adverse direction is based on the direction that gives the best fit to the specific model.

To identify the best-fit model for each feature, the Akaike information criterion (AIC) will be used[23]. AIC provides an estimation of the relative quality of models for a given set of data by balancing the complexity of the model with how well the model fits the data. The model with the lowest AIC will be selected as the final model and used to calculate the probe set BMD, BMD lower confidence limit (BMD_L), and BMD upper bound (BMD_U)[24]. For some features, the best-fit model has an estimated BMD that is well below (i.e., three-fold or more) the lowest positive dose in the study. This case occurs most often with the Hill model, but also can occur with the exponential 4 and 5 models. Due to how the data are reported (median gene-set BMD), some small amount of extreme model extrapolation can be tolerated. If, however, the median BMD for a gene set is derived from a gene with extreme model extrapolation, identifying implausibly low gene-set BMD values is possible. In such cases, the ideal solution is to rerun the study with additional lower dose levels, but resource constraints sometimes preclude reruns. To facilitate resolution of this issue without running an additional study, several approaches have been developed to deal with extreme model extrapolation, many of which are implemented in the BMDExpress software. NTP will flag all BMD values from best-fit models that are estimated to be less than 10-fold the lowest positive dose but allow them to pass through into the gene-set analysis assuming they meet other criteria (as defined in the next section of this document). If these flagged BMD values are identified as the median BMD value for a gene set (described in greater detail in the next section), the gene set will be flagged in the subsequent part of the analysis and reported simply as having a BMD value <10-fold below the lowest positive dose. In limited cases, all model executables will fail parameter convergence (i.e., BMD, BMD_L, or BMD_U). In such cases, the feature will not be considered in the gene-set analysis described below. Note: The US EPA is currently working on methods for model averaging that will obviate the need to select a best-fit model[25]. NTP plans to evaluate model averaging for use in genomic dose-response analysis once the US EPA establishes it as a viable alternative to the best model selection approach.

Justification for approach: The choice of using the parametric models specified in the US EPA BMDS software is based, in part, on validation of the model algorithms for use in regulatory data modeling and their greater simplicity. With the scale of data that NTP will generate, parametric models provide a considerable gain in computational efficiency as compared to alternative approaches[26,27].

NTP will use all continuous models currently available from the US EPA use in the modeling, in part, because a prior hypothesis for the behavior of any given transcript is unknown due to the complexity of the biological response. Further, use of multiple models maximizes the likelihood that a feature with dose-related response will fit the model well enough to be considered further in gene-set analysis.

NTP selected BMD as the potency metric because it is consistent with common practice in regulatory toxicology. The BMD potency metric is used in regulatory toxicology because it

8

provides a model-based determination of the minimum dose level expected to have a significant biological effect. The BMD often occurs between the no observed effect level and the lowest observed effect level, both of which are determined by more traditional, pairwise statistical analysis.

A BMR of 1 SD of the control was selected because it is consistent with US EPA recommendations when modeling continuous data for which an adverse effect level of change is unknown. The US EPA's justification for selection of 1 SD relates to the observation that a 1-SD BMR at control when the data are normally distributed yields an excess risk of approximately 10% for the proportion of individuals less than the 2nd percentile or greater than the 98th percentile of controls. Thus, the 1-SD threshold approximates a 10% change in response consistent with the commonly used BMR used to model dichotomous data.

Prior to data analysis and after data normalization, the counts (in the case of RNA-Seq) or intensities (in the case of microarray) from the measured features are log transformed. Log transformation of data is generally agreed to produce data that approximate a normal distribution with predominantly constant variance[28]; therefore, the data are assumed to exhibit constant variance. As noted above, future versions of the BMDExpress software will implement a test for heteroscedasticity for each feature, and the results will inform the modeling assumption of constant or nonconstant variance.

For best model selection, NTP will use an AIC. This approach is consistent with current standards of practice recommended by the US EPA. Using AIC for model selection helps balance the goodness of fit that favors smaller residual error to a model, but negatively selects for overparameterization (excessive complexity), which helps avoid overfitting.

Determining Gene-Set Level Potencies

Approach: After selecting the best model, the measured features and their associated BMD/BMD$_L$/BMD$_U$ values are parsed into predefined gene sets (the specific gene sets are discussed below). For features to be passed into the gene-set analysis, the best-fit model must: (1) demonstrate convergent BMD, BMD$_L$ and BMD$_U$ values; (2) have a BMD less than the highest positive dose used in the study; (3) not map to more than one gene; (4) have a model fit p-value > 0.1 determined by a likelihood ratio test; and (5) have a BMD$_U$-to-BMD$_L$ ratio <40. All features passing these selection criteria are converted to their corresponding National Center for Biotechnology Information Entrez Gene ID and then parsed into predefined gene sets. Gene sets that contain at least three genes and are at least 5% populated (based on total annotated gene number) with genes from the study are declared "active," and BMD, BMD$_L$ and BMD$_U$ are determined by calculating a median BMD, BMD$_L$, and BMD$_U$ for each "active" gene set. When the median BMD is less than 10-fold the lowest positive dose (i.e., the feature associated with the gene is flagged, as noted above), the BMD for that gene set will be reported as <10-fold the lowest positive dose and the BMD$_L$ and BMD$_U$ will not be reported.

Justification for approach: Features that exhibit no convergent BMDs or BMDs greater than the highest dose are removed because of significant uncertainty in the model, the derived potency, or both. Features that map to more than one gene are removed because the gene the signal represents is uncertain. A fit p-value > 0.1 is a standard threshold the US EPA recommends to ensure adequate goodness of fit of the data to a model. In certain cases, the global goodness-of-fit statistic will filter features exhibiting a clear dose-response due to poor fit of the model to a

9

subset of dose groups. Notably, these features often measure genes considered as biomarker genes (e.g., *CYP3A4* in response to rifampicin in human hepatocytes). Although loss of these features often has limited impact on the gene set-level potency estimates, they undermine the biologist's confidence in and comfort with the proposed approach to modeling. For this reason, future efforts related to a measure of a goodness of fit will seek potentially to replace or combine the log likelihood fit p-value test with an evaluation of fit residuals around the BMD and an R^2 statistic. Finally, a BMD_U-to-BMD_L threshold <40 removes features with high uncertainty in their fit to the model, which is often related to noisy data that should not be considered in the gene-set analysis.

The thresholds for calling a gene set "active" have been subject to significant debate, as most notably indicated by the many different thresholds used in the peer-reviewed literature[29]. Selection of the three-gene and 5% populated thresholds have been based on expert opinion to represent general minimum thresholds required to consider a gene set to be altered by chemical treatment (i.e., active). Further, three is the minimum number of genes to determine a median value. Future efforts will consider the use of resampling-based statistical tests for gene set-level enrichment to align analysis protocols related to genomic dose-response analysis more closely with best practices used in analysis of functional genomic studies. Importantly, even with the additional filters based on enrichment statistics, hard thresholds for identifying "active" gene sets likely will remain necessary.

The selection of median BMD, BMD_L, and BMD_U values as representative gene-set potency values is based on estimating the central tendency of the gene BMDs in a gene set. This relatively simple approach to estimating gene set-level potency has been effective in limited empirical assessments. In future efforts, NTP will explore the use of bootstrapping-based identification of confidence intervals of the median gene-set BMD. In addition, NTP will evaluate the use of predefined, absolute-rank BMD values for reporting gene-set BMDs (i.e., the 2nd or 3rd ranked BMD based on potency). The focus of such an evaluation will be to determine if these alternative methods for estimating gene-set BMD values yield more accurate estimates of the most sensitive apical BMDs and the reproducibility of gene-set BMDs between replicate studies.

Addressing the Study Design

Approach: Two paradigms are related to study design in toxicology studies[30]. Paradigm 1 is the standard study design prescribed by the toxicological testing guidelines of the Organisation for Economic Co-operation and Development, for which the goal is to maximize the power for performing pairwise comparisons between dose groups to identify a no observed effect level[31]. This design is highly dependent on prior knowledge of the biological potency, particularly if applying dose-response, model-based approaches to the data analysis, such as those proposed herein. In paradigm 2, the study design focuses on creating a data set on which dose-response modeling is performed on the entire data set[16]. The goal of this type of analysis is to identify a model-based estimate of the minimal dose that produces an effect (i.e., a BMD). The number of biological replicates for paradigm 2 is less than for paradigm 1, and more dose levels are employed. NTP will use paradigm 2 for in vivo and in vitro GDRS. Note: NTP's initial efforts in the area of in vivo genomic dose-response used paradigm 1 to allow for comparison with traditional measures of toxicity that often requires well-powered pairwise statistical methods, particularly in the case of dichotomous data such as pathology. Thus, many of the early studies

10

reported by NTP have fewer dose levels (≤7) than proposed herein. All future studies, however, will use paradigm 2.

In addition to the two paradigms described above, several critical study design parameters require discussion, including the species/sex of the model system, age of the animals at study start (in vivo), duration of exposure, organ (in vivo studies) or cell type (in vitro studies) selection for genomic analysis, and high-dose selection.

For in vivo studies, NTP proposes the following study parameters:

- Using 6- to 8-week-old male or female rats. Prior to the main study, a dose range finding study will be performed, and the sex exhibiting greater dose-related sensitivity will be used in the subsequent genomic dose-response study. This will allow for consideration of both sexes and lead to the identification of more sensitive/protective gene set-level BMD values.

- Setting exposure duration to 5 days (i.e., 5 repeated doses, 1 per day, followed by euthanasia 24 hours after the last dose). Duration of exposure might be extended or paired kinetic studies to allow for dose adjustment performed for chemicals that do not achieve steady state after 5 days of dosing. In silico-based modeling will be employed to determine if confounding toxicokinetic parameters need to be considered when determining duration of exposure.

- Selecting organs/tissues for evaluation based on route of proposed exposure and review of the literature with a specific focus on target organs from guideline toxicity studies of the test article or structurally related test articles. In addition, liver will be evaluated in all studies, as it is commonly affected by chemical challenge and often serves as a biosensor of systemic toxicological effects.

- Setting the highest dose for a test article based on determining the 5-day maximum tolerated dose (MTD; i.e., the highest dose that produces less than 20% decrease in body weight gain after 5 days of repeated dosing).

For in vitro studies, NTP proposes the following study parameters:

- Using human cell lines/types with sex as male or female depending on availability.

- Setting exposure duration based on expert review and results from range-finding studies.

- Selecting cell type based on several variables including culturability in 384-well format; patterns of use in the field (i.e., cell types that are more commonly used by the testing community will be given higher priority); and representation of common tissues of concern (i.e., organotypic representation of in vivo target tissues such as liver, e.g., HepaRG).

- Setting the highest dose for the test article to target the lethal concentration 20 (LC20; 20% reduction in cell viability relative to control). LC20 values will be determined by a range-finding study. Alternatively, some chemicals will not reach an LC20, and the high dose will be set at the solubility limit.

Justification for approach: With most test articles, prior knowledge of their biological (toxicogenomic) potency is limited; therefore, NTP will employ a broad dose range to capture the entire dose-response space adequately for all measured features. This study design is

11

consistent with recommendations for the design of studies for which the goal is deriving a BMD[18].

For in vivo study parameters, 6- to 8-week-old male or female rats were chosen, in part, to be congruent with subchronic toxicological assessments and existing toxicogenomic databases such as DrugMatrix[32] and Open TG-Gates[33], which commonly use this strain in this age range. Rats in this age range grow rapidly and, therefore, often are more susceptible to toxicological challenge. Legacy toxicogenomics data are often used for interpreting results obtained from new test articles.

The duration of exposure selected was based on findings by Thomas et al.[19], which demonstrated that 5 days of exposure is sufficient for non-bioaccumulative chemicals to achieve a genomic point of departure similar to that observed for apical endpoints such as cancer. As noted above, if prestudy kinetic predictions suggest the test article has bioaccumulative properties, the study duration could be extended to allow for achievement of steady-state levels. Alternatively, paired pharmacokinetic studies will be performed to allow kinetic-based adjustment of biological potency estimates[34]. Future studies will consider the use of an earlier time point to evaluate the temporal effects on biology and begin to associate early findings with later effects to build the necessary mechanistic bridges likely required for use of genomic dose-response results in a formal risk assessment setting.

Selecting organs/tissues for gene expression studies based on anticipated target organs is common practice in targeted toxicological assessments. Further, analyzing effects on the liver for every test article, particularly if administered by the oral route, is justified primarily because liver is the most frequently affected toxicological target[35] and effects in this organ often are used as the basis for selecting a point of departure when determining safe exposure levels. Second, liver responds to significant disease-related changes in other organ systems and, therefore, often can serve as a sensor for systemic toxicological effects[36].

Using a 5-day MTD as the highest dose helps ensure a clear toxicogenomic response at the top dose level that can be modeled effectively into lower dose levels. A strong response at the top dose is critical because, without a clear toxicogenomic response, identification of features that are responding is uncertain and leads to inaccurate estimates of potency.

For in vitro study parameters, the most critical decision point is likely the cell types to use. NTP has chosen to use primary and organotypic cell systems as opposed to transformed cells because we believe the data derived from primary and organotypic cultures will have greater qualitative and quantitative biological relevance to the organs/tissues we hope to model[37]. Further, using organotypic cell types enables evaluation of known, prototypical, expression-related changes that have been observed in intact organisms, thereby providing a level of model validation[37]. NTP is not naïve to the challenges associated with primary and organotypic cultures, such as reproducibility due to genetic variation, sensitivity to culture environment, and complexity[38]. Prior to using systems, issues of reproducibility will be assessed through in-depth model characterization, replication studies, development of detailed protocols, and use of culture systems that promote stable cellular phenotypes such as 3D spheroids[39]. Similar to the MTD-based selection of the highest dose for in vivo studies, the highest dose in in vitro studies is meant to challenge the cells significantly to obtain a clear toxicogenomic response at the top dose level that can be modeled into lower dose levels.

Addressing Biological Interpretation

Approach: A major challenge for interpreting toxicogenomics data relates to linking the observations to higher-order phenotypes, such as pathology and other endpoints typically associated with adverse effects[40]. NTP acknowledges that such a linkage between genomic changes and more standard metrics of adversity likely will be necessary to gain acceptance in the more formal risk assessment process. Building a collection of phenotypically linked gene sets, however, is a major undertaking that should be accomplished through a community effort with oversight from toxicologists familiar with pathology and genomics. Many curated gene sets are publicly available, such as the KEGG (Kyoto Encyclopedia of Genes and Genomes)[41] and Gene Ontology resources[17]. Although these databases have several shortcomings, such as redundancy in certain areas of biology (e.g., p53 signaling) and poor coverage in uncharacterized biological space, the existing gene sets collectively cover a sizable portion of characterized biological space and likely are adequate to describe many of the biological processes of toxicological concern. The challenge with the existing curated gene sets is the lack of a formal association with toxicological processes.

NTP is focused on identifying the most potent gene set-level biological effect in the genomic dose-response study, and the lack of linkage to adverse effects therefore is not critical. As noted above, to advance the results of these studies beyond screening-level assessments of potency, such linkage likely will become essential. Considering NTP's goal of identifying biological potency and recognizing the challenges noted above in creating a toxicity-associated collection of gene sets, we will use two large curated, public gene sets for analysis of genomic dose-response data: the MSigDB C2[42] and the Gene Ontology Biological Processes[17]. The C2 gene set is a collection of curated pathways (i.e., structured relationships between the genes are defined, gene sets represent cell signaling processes) from well-known public resources such as KEGG[41], Reactome[43], and Biocarta[44], in addition to published chemical and genetic perturbation signatures. The gene sets in the Gene Ontology Biological Processes are expert curated and unstructured and are recognized as contributing a vast array of biological processes (e.g., fatty acid metabolism, cell migration). Often, these gene sets are highly redundant, which might be advantageous, as the identification of multiple, related gene sets can increase confidence that treatment is altering biological space. Further, the likelihood also increases that the most sensitive combination of genes in a given biological space will be represented and therefore will identify the lowest possible, most protective gene-set BMD in a given biological space. Use of these gene sets will enable development and evaluation of genomic dose-response in the manner proposed herein. The comprehensive nature of the gene sets will ensure broad coverage of biological space and, therefore, sensitive BEPODs. Future efforts will focus on using an adverse outcome pathway framework to start mapping gene sets from the Gene Ontology to key events associated with toxicity. Once established, these linkages can be used to draw more formal conclusions about the toxicological significance of findings from GDRS.

NTP is best known for its efforts related to hazard identification using guideline testing methods. Although these efforts protect public health, they also can influence commerce significantly. Thus, a high degree of certainty must be associated with NTP's hazard identification process. In some cases, gene sets from MSigDB or Gene Ontology that might be identified as "active" in GDRS have names and descriptions that suggest a hazard (e.g., Response to DNA Damage, in the case of genotoxity). Although in certain cases effects on such a gene set might be related to a hazard such as genotoxicity, it is not always the case when determined by guideline testing

13

methods (e.g., Ames assay). Several reasons for this include how the gene sets were curated (i.e., they might not be meant specifically for interpreting gene expression data), the direction of response in the gene set indicates different effects, and the diversity of response to genes in the set is based on the biological setting (i.e., different organs or cell types). Although unlikely, it is also possible that the guideline testing method is wrong in certain cases and the gene set could be correctly indicating a hazard. Due to the uncertainty and to avoid potential confusion, NTP will limit the interpretation of the findings to descriptions of the gene sets the curators provide. In addition, any reporting of genomic dose studies will include clarifying statements indicating that the studies should not be used for formal hazard identification/characterization. This is, of course, subject to change as consensus is formed on how to state the confidence in the association between gene sets and key events related to adverse effects (e.g., more formal adverse outcome pathway-based linkage, as noted above).

In addition to the limited qualitative interpretation of the genomic dose-response data, NTP proposes to interpret the results quantitatively by identifying the BMD median from the most sensitive gene set (i.e., the BMD from the gene set with the lowest BMD). The BMD median from this gene set is described as the dose at which a test article begins to demonstrate an effect on a biological system or a BEPOD, as described above.

Justification for approach: Use of the MSigDB C2 and Gene Ontology Biological Process sets will provide broad coverage of biological space and capture sensitive gene-set BMD values. Although formulating adverse effect-associated gene sets will be pursued, requiring this to advance the approach presented here will hamper progress toward developing a biological potency-focused assessment unnecessarily.

NTP wishes to avoid confusing users about the goal of performing GDRS, which is first and foremost to identify biological potency. The names and descriptions of certain gene sets might be misleading regarding their association with adverse toxicological effects. To avoid the challenge of interpretation in the context of hazard, standard language will be used in the introduction of each report to make clear that GDRS results should not be used to make formal hazard conclusions at this time.

Using the BMD median of the most sensitive gene set to determine the most sensitive biological effect level is based on the observation from in vivo GDRS that this approach leads to identification of the biological effect level that approximates the point of departure from guideline toxicology studies[19].

References

1. Collins FS, Gray GM, Bucher JR. Transforming environmental health protection. Science. 2008; 319(5865):906-907. http://dx.doi.org/10.1126/science.1154619

2. Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, Devito MJ, Cote I, Rowlands JC, Whelan MP, Hays SM, Andersen ME et al. Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework. Toxicol Sci. 2013; 136(1):4-18. http://dx.doi.org/10.1093/toxsci/kft178

3. Thomas RS, Allen BC, Nong A, Yang L, Bermudez E, Clewell HJ, 3rd, Andersen ME A method to integrate benchmark dose estimates with genomic data to assess the functional effects of chemical exposure. Toxicol Sci. 2007; 98(1):240-248. http://dx.doi.org/10.1093/toxsci/kfm092

4. Bhat VS, Hester SD, Nesnow S, Eastmond DA Concordance of transcriptional and apical benchmark dose levels for Conazole-induced liver effects in mice. Toxicological Sciences. 2013; 136(1):205-215. http://dx.doi.org/10.1093/toxsci/kft182

5. Dunnick JK, Shockley KR, Morgan DL, Brix A, Travlos GS, Gerrish K, Michael Sanders J, Ton TV, Pandiri AR Hepatic transcriptomic alterations for N,N-dimethyl-p-toluidine (DMPT) and p-toluidine after 5-day exposure in rats. Arch Toxicol. 2017; 91(4):1685-1696. http://dx.doi.org/10.1007/s00204-016-1831-7

6. Fader KA, Nault R, Zhang C, Kumagai K, Harkema JR, Zacharewski TR 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD)-elicited effects on bile acid homeostasis: Alterations in biosynthesis, enterohepatic circulation, and microbial metabolism. Sci Rep. 2017; 7. http://dx.doi.org/10.1038/s41598-017-05656-8

7. Kawamoto T, Ito Y, Morita O, Honda H Mechanism-based risk assessment strategy for drug-induced cholestasis using the transcriptional benchmark dose derived by toxicogenomics. Journal of Toxicological Sciences. 2017; 42(4):427-436. http://dx.doi.org/10.2131/jts.42.427

8. Rager JE, Ring CL, Fry RC, Suh M, Proctor DM, Haws LC, Harris MA, Thompson CM High-throughput screening data interpretation in the context of in vivo transcriptomic responses to oral Cr(VI) exposure. Toxicological Sciences. 2017; 158(1):199-212. http://dx.doi.org/10.1093/toxsci/kfx085

9. Robinson JF, Tonk ECM, Verhoef A, Piersma AH Triazole induced concentration-related gene signatures in rat whole embryo culture. Reproductive Toxicology. 2012; 34(2):275-283. http://dx.doi.org/10.1016/j.reprotox.2012.05.088

10. Rowlands JC, Budinsky R, Gollapudi B, Black MB, Wolfinger RD, Cukovic D, Dombkowski A, Thompson CM, Urban JD, Thomas RS A genomics-based analysis of relative potencies of dioxin-like compounds in primary rat hepatocytes. Toxicol Sci. 2013; 136(2):595-604. http://dx.doi.org/10.1093/toxsci/kft203

11. Thompson CM, Rager JE, Suh M, Ring CL, Proctor DM, Haws LC, Fry RC, Harris MA Transcriptomic responses in the oral cavity of F344 rats and B6C3F1 mice following exposure to

15

Cr(VI): implications for risk assessment. Environmental and Molecular Mutagenesis. 2016; 57(9):706-716. http://dx.doi.org/10.1002/em.22064

12. Yang Y, Maxwell A, Zhang XW, Wang N, Perkins EJ, Zhang CY, Gong P Differential reconstructed gene interaction networks for deriving toxicity threshold in chemical risk assessment. Bmc Bioinformatics. 2013; 14. http://dx.doi.org/10.1186/1471-2105-14-S14-S3

13. Dean JL, Zhao QJ, Lambert JC, Hawkins BS, Thomas RS, Wesselkamper SC Editor's highlight: Application of gene set enrichment analysis for identification of chemically induced, biologically relevant transcriptomic networks and potential utilization in human health risk assessment. Toxicol Sci. 2017; 157(1):85-99. http://dx.doi.org/10.1093/toxsci/kfx021

14. Jaroch K, Jaroch A, Bojko B Cell cultures in drug discovery and development: The need of reliable in vitro-in vivo extrapolation for pharmacodynamics and pharmacokinetics assessment. J Pharm Biomed Anal. 2018; 147:297-312. http://dx.doi.org/10.1016/j.jpba.2017.07.023

15. Yang L, Allen BC, Thomas RS BMDExpress: a software tool for the benchmark dose analyses of genomic data. BMC Genomics. 2007; 8:387. http://dx.doi.org/10.1186/1471-2164-8-387

16. Crump KS A new method for determining allowable daily intakes. Fundam Appl Toxicol. 1984; 4(5):854-871. http://dx.doi.org/10.1016/0272-0590(84)90107-6

17. Gene Ontology Consortium. The Gene Ontology project in 2008. Nucleic Acids Res. 2008; 36(Database issue):D440-444. http://dx.doi.org/10.1093/nar/gkm883

18. Slob W, Moerbeek M, Rauniomaa E, Piersma AH A statistical evaluation of toxicity study designs for the estimation of the benchmark dose in continuous endpoints. Toxicol Sci. 2005; 84(1):167-185. Toxicol Sci. http://dx.doi.org/10.1093/toxsci/kfi004

19. Thomas RS, Wesselkamper SC, Wang NC, Zhao QJ, Petersen DD, Lambert JC, Cote I, Yang L, Healy E, Black MB et al. Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. Toxicol Sci. 2013; 134(1):180-194. http://dx.doi.org/10.1093/toxsci/kft094

20. Williams DA A test for differences between treatment means when several dose levels are compared with a zero dose control. Biometrics. 1971; 27(1):103-117. http://dx.doi.org/10.2307/2528930

21. Yeakley JM, Shepard PJ, Goyena DE, VanSteenhouse HC, McComb JD, Seligmann BE A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling. PLoS One. 2017; 12(5):e0178302. http://dx.doi.org/10.1371/journal.pone.0178302

22. MAQC Consortium The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol. 2006; 24(9):1151-1161. Nat Biotechnol. http://dx.doi.org/10.1038/nbt1239

23. Akaike H. Information theory and an extension of the maximum likelihood principle. Second International Symposium on Information Theory. 1973. p. 267-281.

24. Crump KS Calculation of benchmark doses from continuous data. Risk Analysis. 1995; 15(1):79-89. http://dx.doi.org/10.1111/j.1539-6924.1995.tb00095.x

25. Shao K, Gift JS Model uncertainty and Bayesian model averaged benchmark dose estimation for continuous data. Risk Anal. 2014; 34(1):101-120. http://dx.doi.org/10.1111/risa.12078

26. Shockley KR Estimating potency in high-throughput screening experiments by maximizing the rate of change in weighted Shannon entropy. Sci Rep. 2016; 6:27897. Sci Rep. http://dx.doi.org/10.1038/srep27897

27. Budtz-Jorgensen E, Keiding N, Grandjean P Benchmark dose calculation from epidemiological data. Biometrics. 2001; 57(3):698-706. Biometrics. http://dx.doi.org/10.1111/j.0006-341X.2001.00698.x

28. Hoyle DC, Rattray M, Jupp R, Brass A Making sense of microarray data distributions. Bioinformatics. 2002; 18(4):576-584. Bioinformatics. http://dx.doi.org/10.1093/bioinformatics/18.4.576

29. Farmahin R, Williams A, Kuo B, Chepelev NL, Thomas RS, Barton-Maclaren TS, Curran IH, Nong A, Wade MG, Yauk CL Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment. Arch Toxicol. 2017; 91(5):2045-2065. http://dx.doi.org/10.1007/s00204-016-1886-5

30. Holland-Letz T, Kopp-Schneider A Optimal experimental designs for dose-response studies with continuous endpoints. Arch Toxicol. 2015; 89(11):2059-2068. Arch Toxicol. http://dx.doi.org/10.1007/s00204-014-1335-2

31. OECD guidance document for the design and conduct of chronic toxicity and carcinogenicity studies, supporting TG 451, 452 and 453. https://www.oecd.org/chemicalsafety/testing/45335229.pdf.

32. Ganter B, Tugendreich S, Pearson CI, Ayanoglu E, Baumhueter S, Bostian KA, Brady L, Browne LJ, Calvin JT, Day GJ et al. Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. J Biotechnol. 2005; 119(3):219-244. http://dx.doi.org/10.1016/j.jbiotec.2005.03.022

33. Igarashi Y, Nakatsu N, Yamashita T, Ono A, Ohno Y, Urushidani T, Yamada H Open TG-GATEs: a large-scale toxicogenomics database. Nucleic Acids Res. 2015; 43(D1):D921-927. http://dx.doi.org/10.1093/nar/gku955

34. Reichard JF, Maier MA, Naumann BD, Pecquet AM, Pfister T, Sandhu R, Sargent EV, Streeter AJ, Weideman PA Toxicokinetic and toxicodynamic considerations when deriving health-based exposure limits for pharmaceuticals. Regul Toxicol Pharmacol. 2016; 79 Suppl 1:S67-78. http://dx.doi.org/10.1016/j.yrtph.2016.05.027

35. Swirsky Gold L, Manley N, Slone T, Ward J Compendium of chemical carcinogens by target organ: results of chronic bioassays in rats, mice, hamsters, dogs, and monkeys. Toxicol Pathol. 2001; 29(6):639-652. http://dx.doi.org/10.1080/019262301753385979

36. Edwards L, Wanless IR Mechanisms of liver involvement in systemic disease. Best Pract Res Clin Gastroenterol. 2013; 27(4):471-483. Best Pract Res Clin Gastroenterol. http://dx.doi.org/10.1016/j.bpg.2013.08.002

37. Gerets HH, Tilmant K, Gerin B, Chanteux H, Depelchin BO, Dhalluin S, Atienzar FA Characterization of primary human hepatocytes, HepG2 cells, and HepaRG cells at the mRNA level and CYP activity in response to inducers and their predictivity for the detection of human hepatotoxins. Cell Biol Toxicol. 2012; 28(2):69-87. Cell Biol Toxicol. http://dx.doi.org/10.1007/s10565-011-9208-4

38. Fatehullah A, Tan SH, Barker N Organoids as an in vitro model of human development and disease. Nat Cell Biol. 2016; 18(3):246-254. http://dx.doi.org/10.1038/ncb3312

39. Ramaiahgari SC, Waidyanatha S, Dixon D, DeVito MJ, Paules RS, Ferguson SS Three-dimensional (3D) HepaRG spheroid model with physiologically relevant xenobiotic metabolism competence and hepatocyte functionality for liver toxicity screening. Toxicol Sci. 2017; 160(1):189-190. http://dx.doi.org/10.1093/toxsci/kfx194

40. Paules R Phenotypic anchoring: linking cause and effect. Environ Health Perspect. 2003; 111(6):A338-339. Environ Health Perspect. http://dx.doi.org/10.1289/ehp.111-a338

41. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017; 45(D1):D353-D361. Nucleic Acids Res. http://dx.doi.org/10.1093/nar/gkw1092

42. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011; 27(12):1739-1740. http://dx.doi.org/10.1093/bioinformatics/btr260

43. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B et al. The Reactome pathway Knowledgebase. Nucleic Acids Res. 2017; 44(D1):D481-D487. https://doi.org/10.1093/nar/gkv1351

44. Biocarta Pathways. http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways.

Appendix A   Comparison of BEPOD to Apical POD
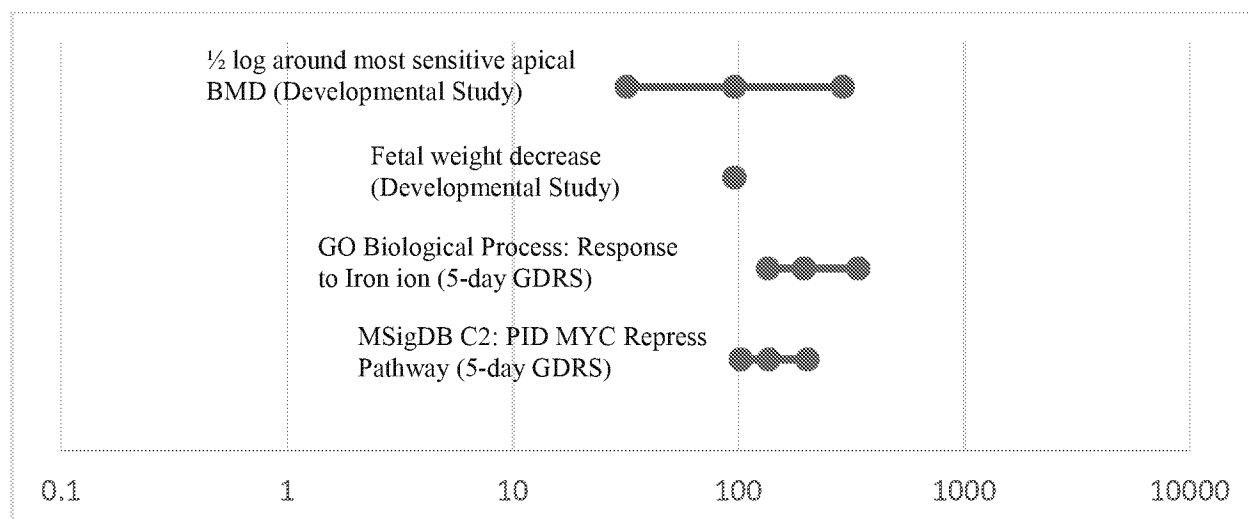
Figures

A.1 Introduction

This appendix contains the results of an analysis using the revised National Toxicology Program (NTP) genomic dose-response study (GDRS) analysis pipeline of in vivo genomic data from rats. The analysis compares the biological effect point of departure (BEPOD) benchmark dose (BMD) values with the most sensitive apical BMD identified from guideline toxicity studies carried out with the same strain of rat used in the genomic studies. The guideline toxicity studies used to identify the most sensitive apical BMD range from 28-day subchronic studies to developmental toxicity and 2-year carcinogenicity studies. When the cancer BMD was not the most sensitive BMD, the cancer BMD values are listed in the figure legend for reference.

A.2 Individual Comparisons of Apical POD to BEPOD
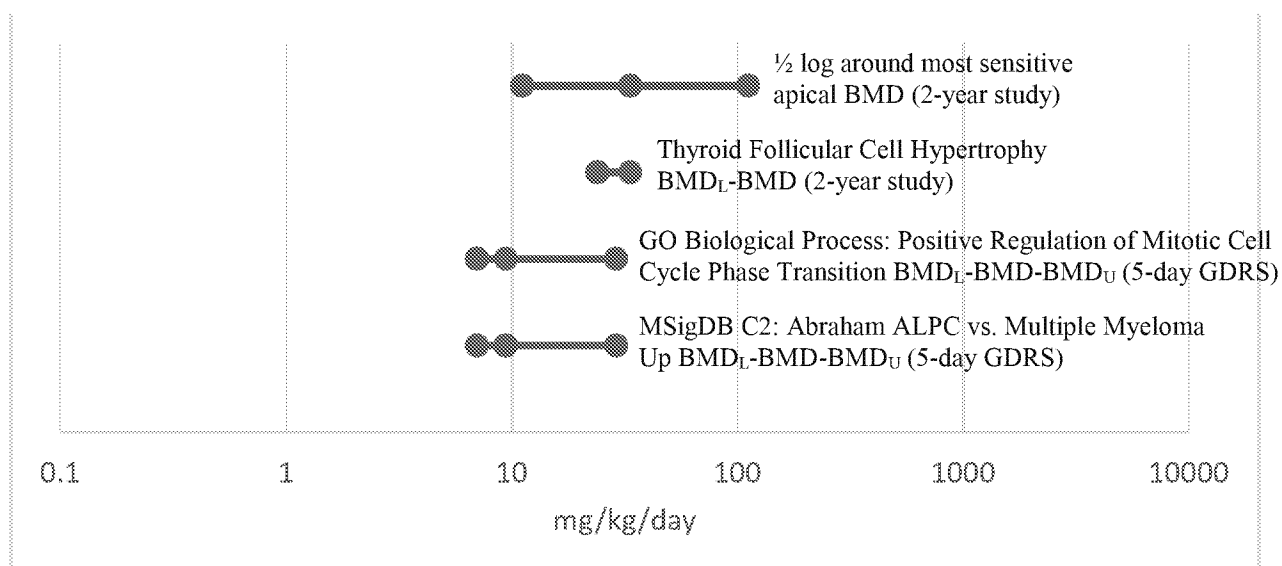
Figure 1. Comparison of the Most Sensitive Apical Endpoint BMD from a 2-Year Cancer Study of *Ginkgo biloba* Extract (GBE) to the Most Sensitive Gene-Set BMD from a 5-Day GDRS of GBE

In a 2-year study, F344/N male rats were exposed to doses of 0, 100, 200, and 1000 mg/kg/day in corn oil[1]. The most sensitive apical endpoint was thyroid follicular cell hypertrophy with BMD10 and BMD10$_L$ values of 33.4 and 23.9 mg/kg/day, respectively (multistage-cancer model on unadjusted incidence counts, 95% confidence limits). A ½ log range above and below the apical BMD of 33.4 mg/kg/day is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. In the 5-day GDRS, F344/N male rats were exposed to doses of 0, 3, 30, 100, 300, and 1000 mg/kg/day in corn oil (Unpublished NTP data). BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD lower confidence limit (BMD$_L$), and BMD upper bound (BMD$_U$) values from the Gene Ontology (GO) Biological Processes[2] and the MSigDB C2 set[3] from gene expression measured in liver of the treated animals. The BMD$_L$-to-BMD$_U$ ranges of the most sensitive gene sets in both cases overlapped with the ½ log range of the most sensitive apical BMD from the 2-year study.
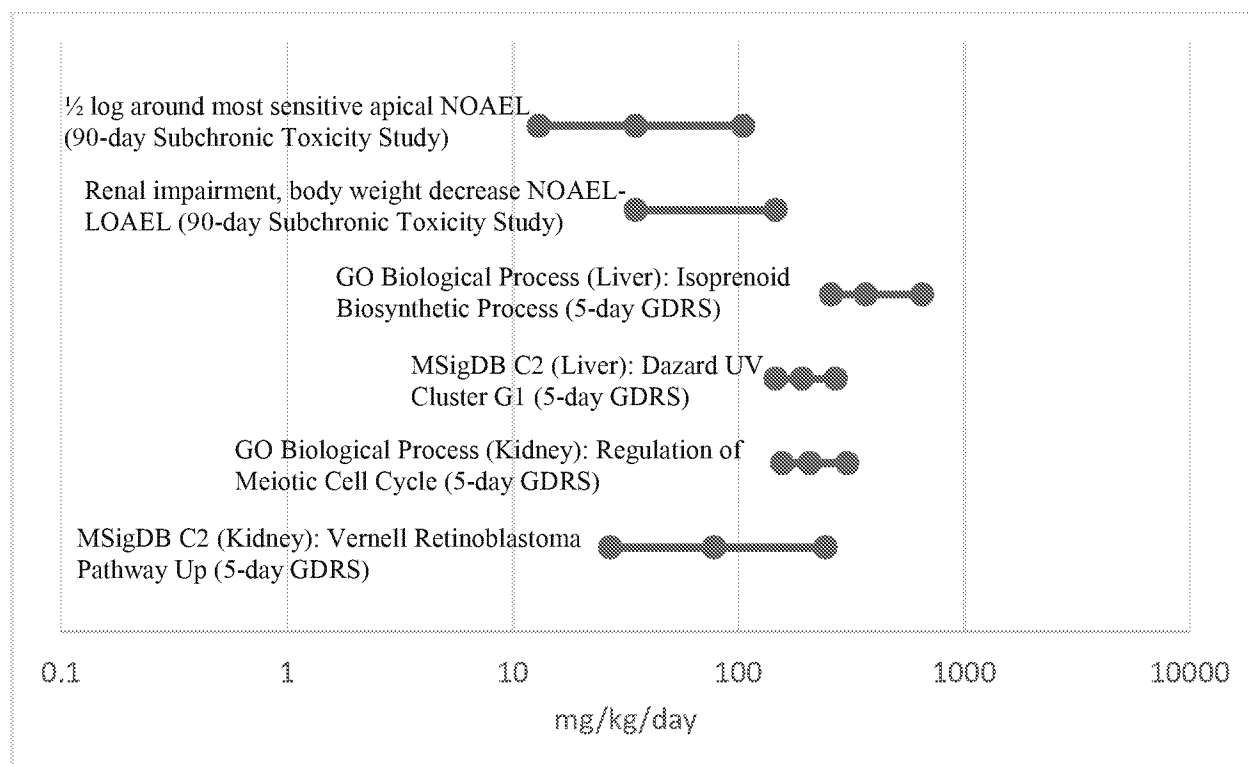
A-3

Figure 2. Comparison of the Most Sensitive Apical Endpoint BMD from a Developmental Toxicity Study of 4-Methylcyclohexanemethanol (MCHM) to the Most Sensitive Gene-Set BMD from a 5-day GDRS of MCHM

In a developmental toxicity study performed by NTP, pregnant Harlan Sprague Dawley rats were exposed to doses of 0, 50, 100, 200, and 400 mg/kg/day in corn oil[4]. The most sensitive apical endpoint identified in the developmental toxicity study was fetal body weight, which exhibited BMD01 and $BMD01_L$ values of 95.9 and 95.6 mg/kg/day, respectively (polynomial, restricted to non-positive coefficients based on relative deviation of fetal weight). A ½ log range above and below the apical BMD of 95.9 mg/kg/day is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. In the 5-day GDRS, male Harlan Sprague Dawley rats were exposed to doses of 0, 0.1, 1, 10, 100, 300, and 500 mg/kg/day in corn oil, and liver gene expression was measured[5]. BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD_L and BMD_U values from the GO Biological Processes[2] and the MSigDB C2 set[3] from gene expression measured in liver of the treated animals. The BMD_L-to-BMD_U ranges of the most sensitive gene sets in both cases overlapped with the ½ log range of the most sensitive apical BMD from the developmental toxicity study of MCHM.

A-4

Figure 3. Comparison of the Most Sensitive Apical Endpoint BMD from a Developmental Toxicity Study of 4-Methylcyclohexanemethanol (MCHM) to the Most Sensitive Gene-Set BMD from a 5-Day GDRS of Crude MCHM

In a developmental toxicity study of MCHM performed by NTP, pregnant Harlan Sprague Dawley rats were exposed to doses of 0, 50, 100, 200, and 400 mg/kg/day in corn oil[4]. The most sensitive apical endpoint identified in the developmental toxicity study was fetal body weight, which exhibited BMD01 and BMD01$_L$ values of 95.9 and 95.6 mg/kg/day, respectively (polynomial, restricted to non-positive coefficients based on relative deviation of fetal weight). A ½ log range above and below the apical BMD of 95.9 mg/kg/day is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. In the 5-day GDRS of crude MCHM (~90% MCHM), male Harlan Sprague Dawley rats were exposed to doses of 0, 0.1, 1, 10, 100, 300, and 500 mg/kg/day in corn oil, and liver gene expression was measured[5]. BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD$_L$ and BMD$_U$ values from the GO Biological Processes[2] and the MSigDB C2 set[3] from gene expression measured in liver of the treated animals. The BMD$_L$-to-BMD$_U$ ranges of the most sensitive gene sets in both cases overlapped with the ½ log range of the most sensitive apical BMD from the developmental toxicity study of MCHM.

Figure 4. Comparison of the Most Sensitive Apical Endpoint BMD from a 90-Day Subchronic Toxicity Study of Propylene Glycol Phenyl Ether (PPH) to the Most Sensitive Gene-Set BMD from a 5-Day GDRS of PPH

In a 90-day subchronic study, male Wistar rats were exposed to doses of 0, 35, 146, and 429 mg/kg/day in drinking water. The most sensitive apical endpoints were mild renal impairment and body weight decreases with NOAEL (no-observed-adverse-effect-level) and LOAEL (lowest-observed-adverse-effect-level) values of 35 and 146 mg/kg/day, respectively[6] (data were not available for BMD modeling). A ½ log range above and below the apical NOAEL of 35 mg/kg/day is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. In the 5-day GDRS, Harlan Sprague Dawley male rats were exposed to doses of 0, 0.1, 1, 10, 100, 500, and 1000 mg/kg/day in corn oil, and liver and kidney gene expression was measured[5]. BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD_L and BMD_U values from the GO Biological Processes[2] and the MSigDB C2 set[3] from gene expression measured in liver and kidney of the treated animals. The BMD_L-to-BMD_U ranges of the most sensitive MSigDB C2 gene set overlapped with the ½ log range of the apical NOAEL from the 90-day subchronic study. The other most sensitive gene-set BMD ranges did not overlap with the apical NOAEL range.
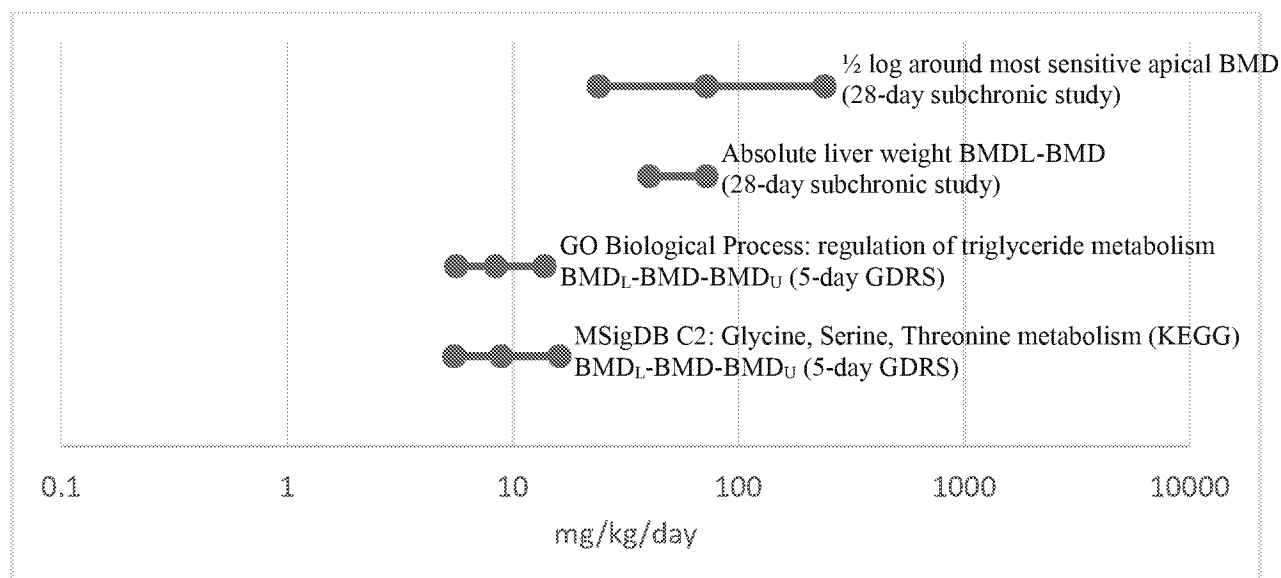
A-6
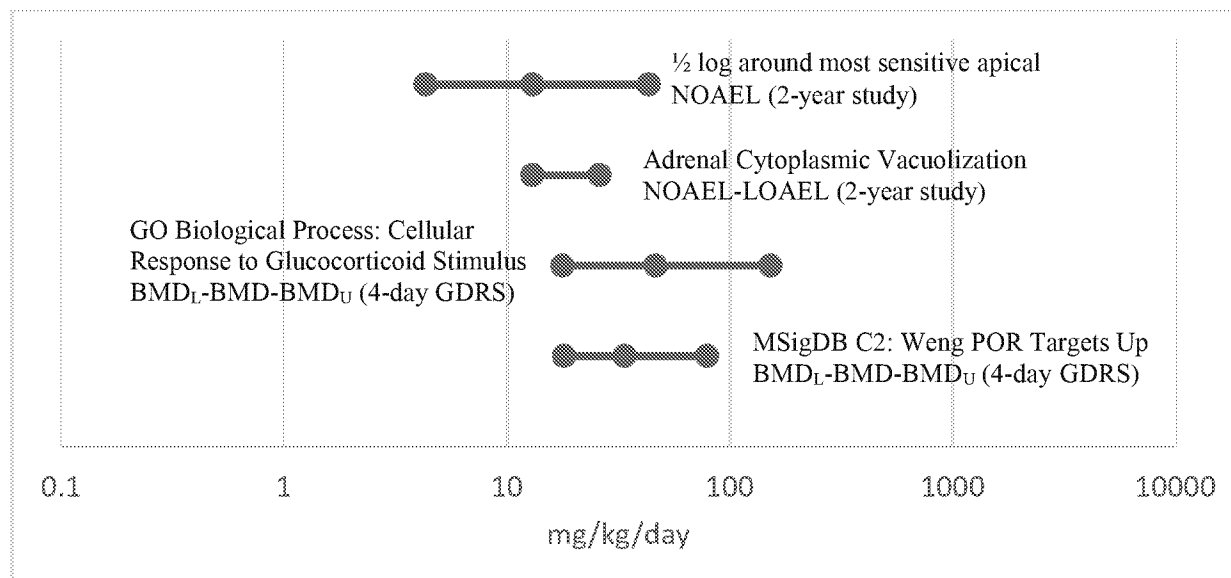
Figure 5. Comparison of the Most Sensitive Apical Endpoint BMD from a 2-Year Cancer Study of Tricresyl Phosphate (TCP) to the Most Sensitive Gene-Set BMD from a 4-Day GDRS of TCP

In the 2-year study, F344/N male rats were exposed to doses of 0, 3, 6, and 13 mg/kg/day in feed (mg/kg/day dose were estimated based on body weight and feed consumption)[7]. The most sensitive apical endpoint was adrenal cortex vacuolization with NOAEL and LOAEL values of 13 and 26 mg/kg/day, respectively (data were inadequate for BMD determination). A ½ log range above and below the apical NOAEL of 13 mg/kg/day is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. In the 4-day GDRS, male Harlan Sprague Dawley rats were exposed to doses of 0, 3, 30, 100, 300, and 1000 mg/kg/day in corn oil, and liver gene expression was measured (NTP unpublished data). BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD_L and BMD_U values from the GO Biological Processes[2] and the MSigDB C2 set[3]. The BMD_L-to-BMD_U ranges of the most sensitive gene sets in both cases overlapped with the ½ log range of the apical NOAEL from the 2-year study.
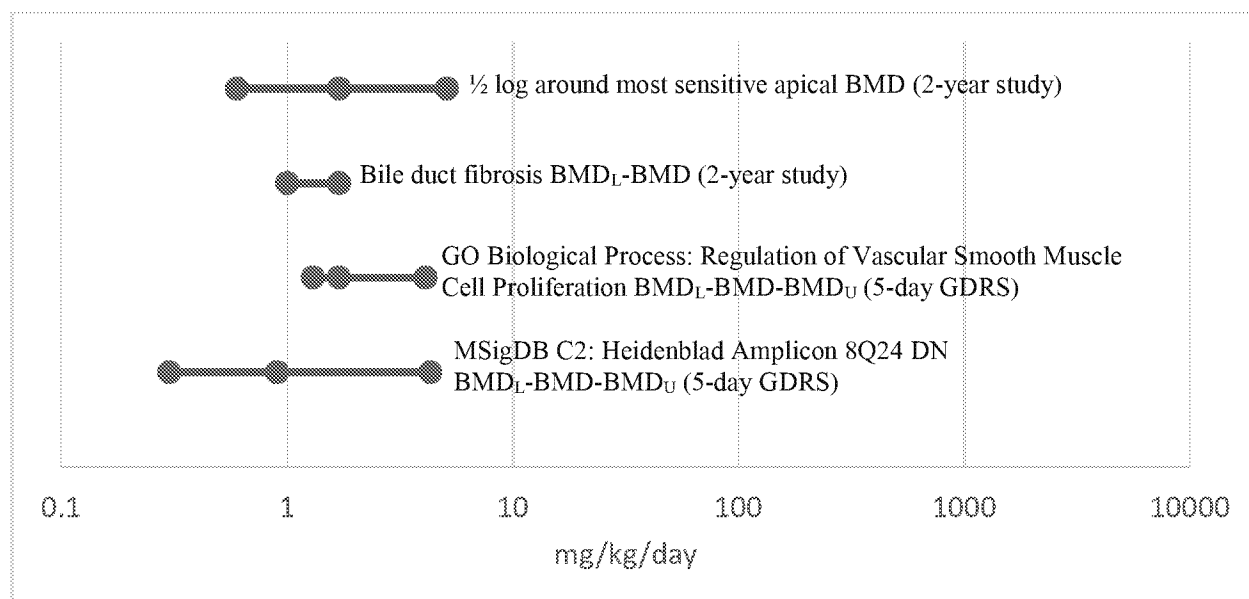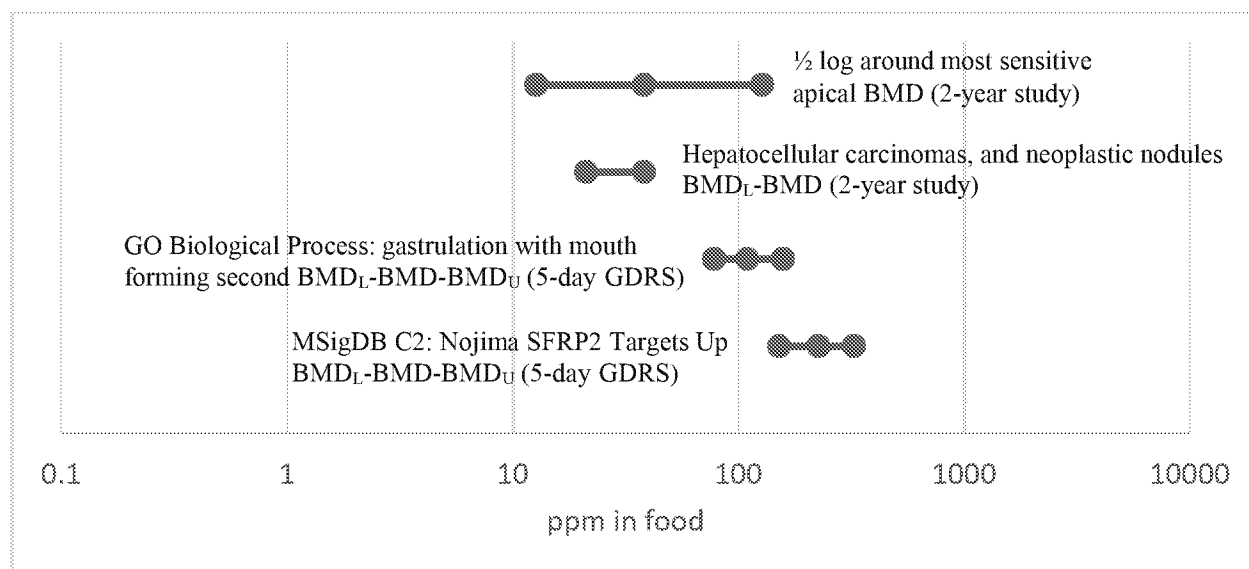
Figure 6. Comparison of the Most Sensitive Apical Endpoint BMD from a 2-Year Cancer Study of N,N-Dimethyl-p-toluidine (DMPT) to the Most Sensitive Gene-Set BMD from a 5-day GDRS of DMPT

In a 2-year study, F344/N male rats were exposed to doses of 0, 6, 20, and 60 mg/kg/day in corn oil[8]. The most sensitive apical endpoint was bile duct fibrosis with BMD10 and BMD10$_L$ values of 1.7 and 1 mg/kg/day, respectively (log-logistic model on incidence, 95% confidence limit). A ½ log range above and below the apical BMD of 1.7 mg/kg/day is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. Not shown on the plot but provided for reference is the hepatocellular carcinoma or adenoma BMD10 and BMD10$_L$ values of 50.1 and 32.6 mg/kg/day, respectively, and the nose adenoma or carcinoma BMD10 and BMD10$_L$ values of 21.1 and 14.0, respectively, that were derived from the 2-year study (log-logistic model on incidence, 95% confidence limit). In the 5-day GDRS, F344/N male rats were exposed to doses of 0, 1, 6, 20, 60, and 120 mg/kg/day in corn oil, and liver gene expression was measured[9]. BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD$_L$ and BMD$_U$ values from the GO Biological Processes[2] and the MSigDB C2 set[3]. The BMD$_L$-to-BMD$_U$ ranges of the most sensitive gene sets in both cases overlapped with the ½ log range of the most sensitive apical BMD from the 2-year study.

½ log around most sensitive apical BMD
(28-day subchronic study)

Absolute liver weight BMDL-BMD
(28-day subchronic study)

GO Biological Process: regulation of triglyceride metabolism
BMD$_L$-BMD-BMD$_U$ (5-day GDRS)

MSigDB C2: Glycine, Serine, Threonine metabolism (KEGG)
BMD$_L$-BMD-BMD$_U$ (5-day GDRS)

0.1   1   10   100   1000   10000

mg/kg/day

Figure 7. Comparison of the Most Sensitive Apical Endpoint BMD from a 28-Day Subchronic Toxicity Study of Bromobenzene (BB) to the Most Sensitive Gene-Set BMD from a 5-Day GDRS of BB

In a 28-day subchronic study, F344/N male rats were exposed to doses of 0, 25, 100, 200, 300, and 400 mg/kg/day in corn oil by gavage[10; 11]. The most sensitive apical endpoint was absolute liver weight with BMD1SD and BMD1SD$_L$ values of 70 and 40.2 mg/kg/day, respectively[11]. A ½ log range above and below the apical BMD of 70 mg/kg/day is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. In the 5-day GDRS, F344 male rats were exposed to doses of 0, 25, 100, 200, 300, and 400 mg/kg/day in corn oil, and liver gene expression was measured[11]. BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD$_L$ and BMD$_U$ values from the GO Biological Processes[2] and the MSigDB C2 set[3]. The BMD$_L$-to-BMD$_U$ ranges of the most sensitive gene sets in both cases were slightly more sensitive than the ½ log range of the most sensitive apical BMD from the 28-day subchronic study.
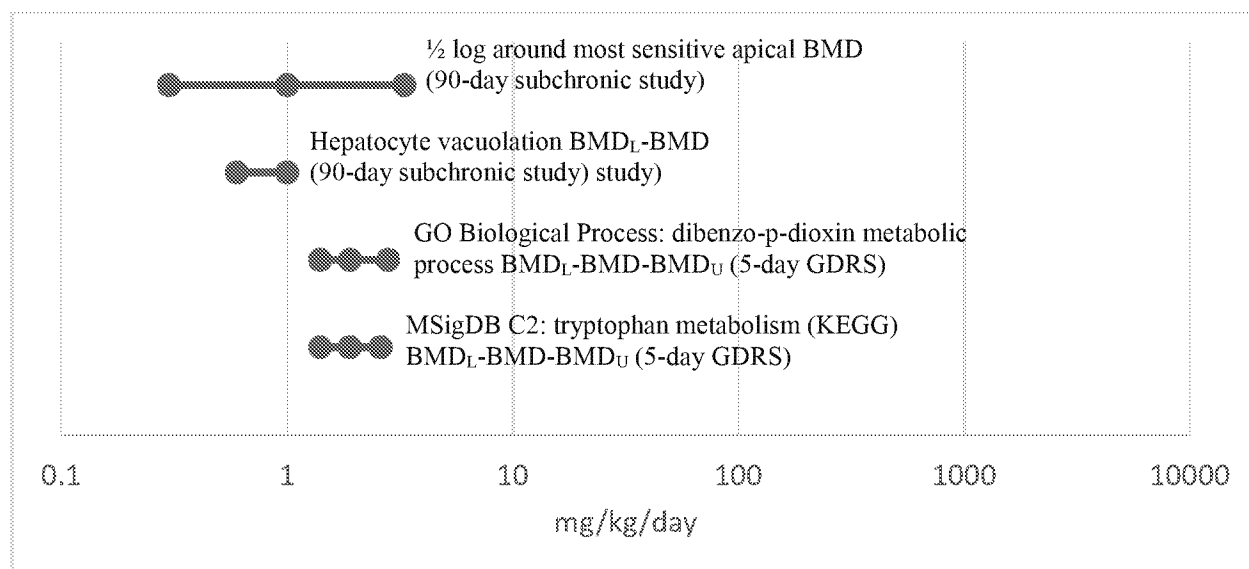
A-9

Figure 8. Comparison of the Most Sensitive Apical Endpoint BMD from a 2-Year Study of Hydroxyazobenzene (HZBZ) to the Most Sensitive Gene-Set BMD from a 5-Day GDRS of HZBZ

In a 2-year study, F344/N male rats were exposed to doses of 0, 80, and 300 ppm in feed[12]. The most sensitive apical endpoint was hepatocellular carcinomas and neoplastic nodules with BMD10 and BMD10$_L$ values of 38.1 and 21.1 ppm, respectively (second-order, multistage-cancer model based on tumor incidence, 95% confidence limits)[11]. A ½ log range above and below the apical BMD of 38.1 ppm is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. In the 5-day GDRS, F344 male rats were exposed to doses of 0, 5, 20, 80, 200, and 300 ppm in feed, and liver gene expression was measured[11; 13]. BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD$_L$ and BMD$_U$ values from the GO Biological Processes[2] and the MSigDB C2 set[3]. The BMD$_L$-to-BMD$_U$ range of the most sensitive GO Biological Processes overlapped with the ½ log range of the most sensitive apical BMD from the 2-year study. The most sensitive MSigDB C2 set BMD$_L$-to-BMD$_U$ range was slightly less sensitive than the ½ log range of the most sensitive apical BMD from the 2-year study.
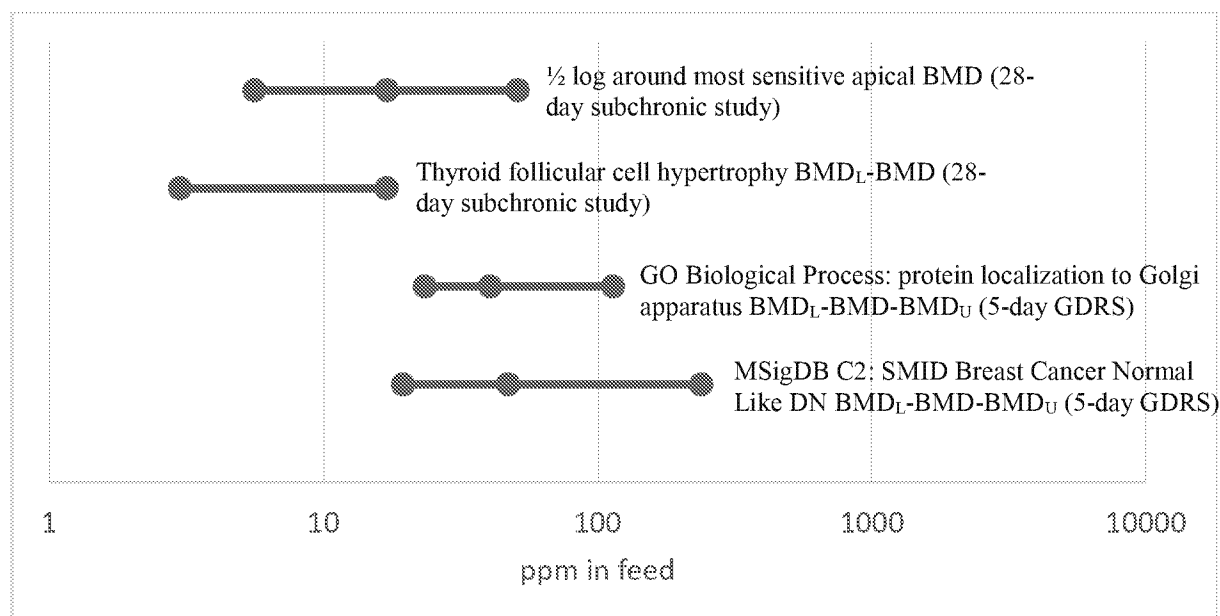
A-10
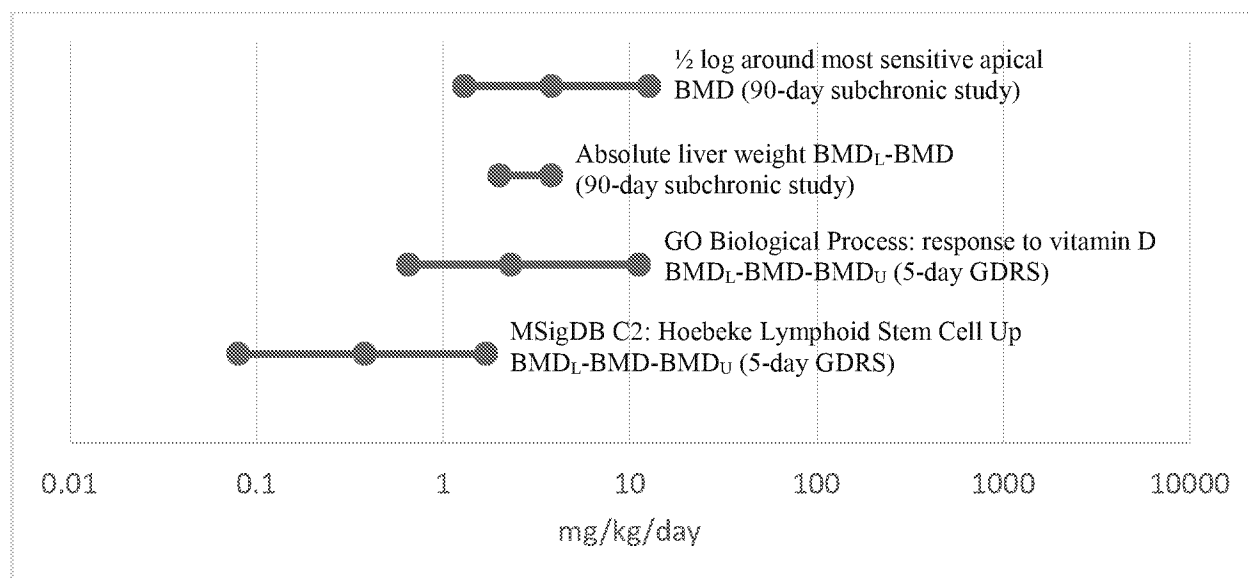
Figure 9. Comparison of the Most Sensitive Apical Endpoint BMD from a 90-Day Subchronic Study of 1,2,4-Tribromobenzene (TRBZ) to the Most Sensitive Gene-Set BMD from a 5-Day GDRS of TRBZ

In a 90-day subchronic study, male Sprague Dawley rats were exposed to doses of 0, 2.5, 5, 10, 25, and 75 mg/kg/day in corn oil[11; 14]. The most sensitive apical endpoint was absolute liver weight with BMD1SD and BMD1SD$_L$ values of 3.8 and 2 mg/kg/day, respectively[11]. A ½ log range above and below the apical BMD of 3.8 mg/kg/day is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. In the 5-day GDRS, male Sprague Dawley rats were exposed to doses of 0, 2.5, 5, 10, 25, and 75 mg/kg/day in corn oil, and liver gene expression was measured[11]. BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD$_L$, and BMD$_U$ values from the GO Biological Processes[2] and the MSigDB C2 set[3]. The BMD$_L$-to-BMD$_U$ ranges of the most sensitive gene sets in both cases overlapped with the ½ log range of the most sensitive apical BMD from the 90-day subchronic study. **Note:** A positive regulation of mRNA metabolic process (from GO Biological Processes), had a BMD < 0.25 mg/kg/day (10-fold <lowest positive dose), and BMD$_L$ and BMD$_U$ values were not reported. To obtain a BMD$_L$-to-BMD$_U$ range for comparison, the next most sensitive process from GO Biological Processes is shown in the plot.
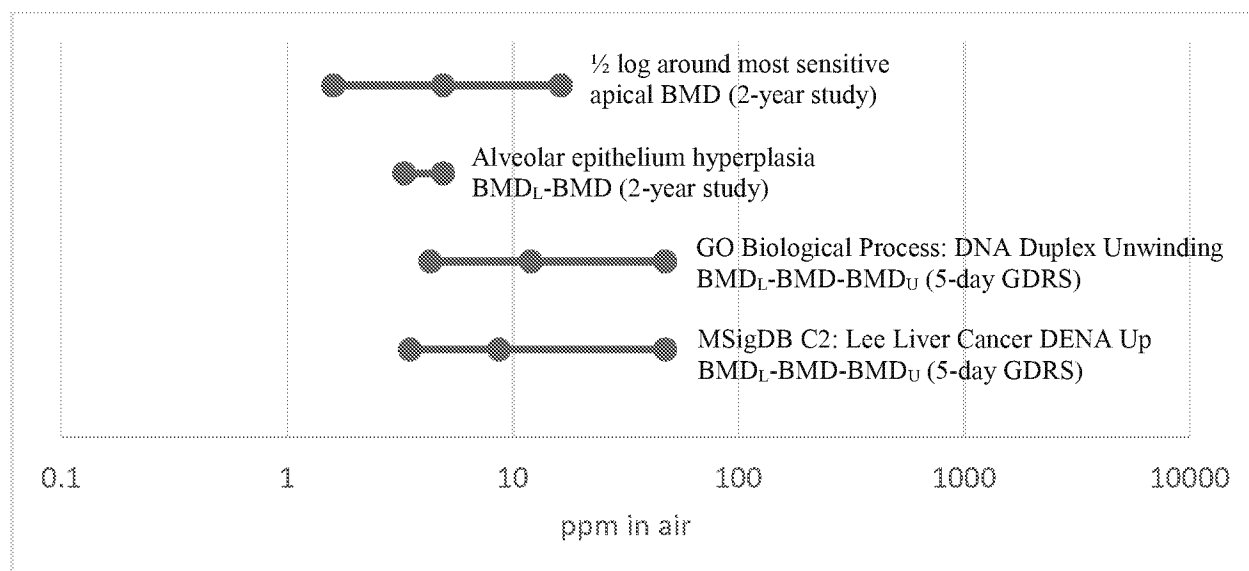
A-11

Figure 10. Comparison of the Most Sensitive Apical Endpoint BMD from a 90-Day Subchronic Study of 2,3,4,6-Tetrachlorophenol (TTCP) to the Most Sensitive Gene-Set BMD from a 5-Day GDRS of TTCP

In a 90-day subchronic study, male Sprague Dawley rats were exposed to doses of 0, 10, 25, 50, 100, and 200 mg/kg/day in corn oil[11; 15]. The most sensitive apical endpoint was hepatocyte vacuolation with BMD10 and BMD10$_L$ values of 1 and 0.6 mg/kg/day, respectively[11]. A ½ log range above and below the apical BMD of 1 mg/kg/day is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. In the 5-day GDRS, male Sprague Dawley rats were exposed to doses of 0, 10, 25, 50, 100, and 200 mg/kg/day in corn oil, and liver gene expression was measured[11]. BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD$_L$, and BMD$_U$ values from the GO Biological Processes[2] and the MSigDB C2 gene sets[3]. The BMD$_L$-to-BMD$_U$ ranges of the most sensitive gene sets in both cases overlapped with the ½ log range of the most sensitive apical BMD from the 90-day subchronic study.

A-12

Figure 11. Comparison of the Most Sensitive Apical Endpoint BMD from a 28-Day Subchronic Study of 4,4′-Methylenebis (N,N-dimethyl) benzenamine (MDMB) to the Most Sensitive Gene-Set BMD from a 5-Day GDRS of MDMB
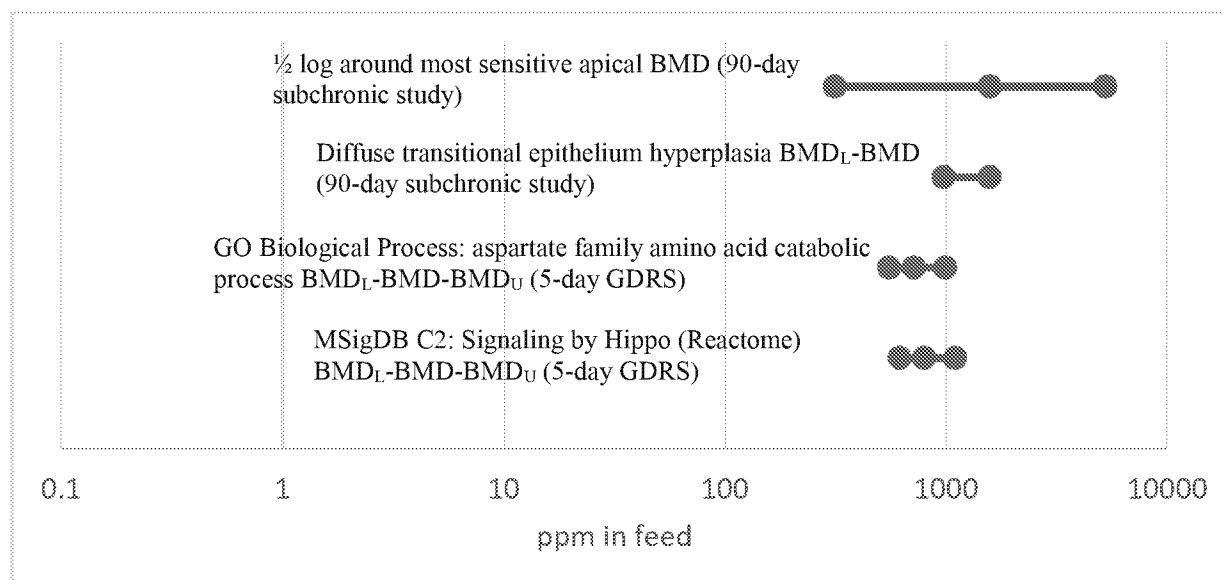
In a 28-day subchronic study, male F344 rats were exposed to doses of 0, 50, 200, 375, 500, and 750 ppm in feed[11; 16]. The most sensitive apical endpoint was thyroid follicular cell hypertrophy with BMD10 and BMD10$_L$ values of 17.0 and 3.1 ppm, respectively[11]. A ½ log range above and below the apical BMD of 17.0 ppm is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. Not shown on the plot is the thyroid follicular cell adenoma and carcinoma BMD and BMD$_L$ values of 381 and 283 ppm, respectively, that were derived from 2-year study findings in the same rat strain[11; 17]. In a 5-day GDRS, male F344 rats were exposed to doses of 0, 50, 200, 375, 500, and 750 ppm in feed, and thyroid gene expression was measured. BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD$_L$ and BMD$_U$ values from the GO Biological Processes[2] and the MSigDB C2 set[3]. The BMD$_L$-to-BMD$_U$ ranges of the most sensitive gene sets in both cases as sensitive as the ½ log range of the most sensitive apical BMD from the 28-day subchronic study.

A-13

Figure 12. Comparison of the Most Sensitive Apical Endpoint BMD from a 2-Year Study of Chloroprene (CHLP) to the Most Sensitive Gene-Set BMD from a 5-Day GDRS of CHLP

In a 2-year study, female F344/N rats were exposed to doses of 0, 12.8, 32, and 80 ppm in air[18]. The most sensitive apical endpoint was alveolar epithelium hyperplasia with BMD10 and BMD10$_L$ values of 4.9 and 3.3 ppm, respectively (log-logistic model, 95% confidence limit). A ½ log range above and below the apical BMD of 17.0 ppm is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. In a 5-day GDRS female F344 rats were exposed to doses of 0, 5, 30, 90, and 200 ppm in air, and lung gene expression was measured[19]. BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD$_L$ and BMD$_U$ values from the GO Biological Processes[2] and the MSigDB C2 set[3]. The BMD-to-BMD$_U$ ranges of the most sensitive gene sets in both cases overlapped with the ½ log range of the most sensitive apical BMD from the 2-year study.

A-14

Figure 13. Comparison of the Most Sensitive Apical Endpoint BMD from a 90-Day Subchronic Study of N-Nitrosodiphenylamine (NDPA) to the Most Sensitive Gene-Set BMD from a 5-Day GDRS of NDPA

In a 90-day subchronic study, female F344 rats were exposed to doses of 0, 250, 1000, 2000, 3000, and 4000 ppm in feed[11; 20]. The most sensitive apical endpoint was diffuse transitional epithelial hyperplasia of the bladder with BMD10 and BMD10$_L$ values of 1567 and 971 ppm, respectively[11]. A ½ log range above and below the apical BMD of 1567 ppm is shown at the top of the plot. The ½ log range is shown to facilitate comparison with the genomic BMD values. Not shown on the plot is the transitional cell carcinoma of the bladder BMD and BMD$_L$ values of 2002 and 1499 ppm, respectively, that were derived from 2-year study findings in the same rat strain[11; 21]. In a 5-day GDRS, male F344 rats were exposed to doses of 0, 250, 1000, 2000, 3000, and 4000 ppm in feed, and bladder gene expression was measured. BMDExpress 2.0 was used to identify the most sensitive gene-set BMD, BMD$_L$, and BMD$_U$ values from the GO Biological Processes[2] and the MSigDB C2 set[3]. The BMD$_L$-to-BMD$_U$ ranges of the most sensitive gene sets in both cases were less sensitive than the ½ log range of the most sensitive apical BMD from the 90-day subchronic study.

A-15

A.3   Global Comparison of POD and BEPOD
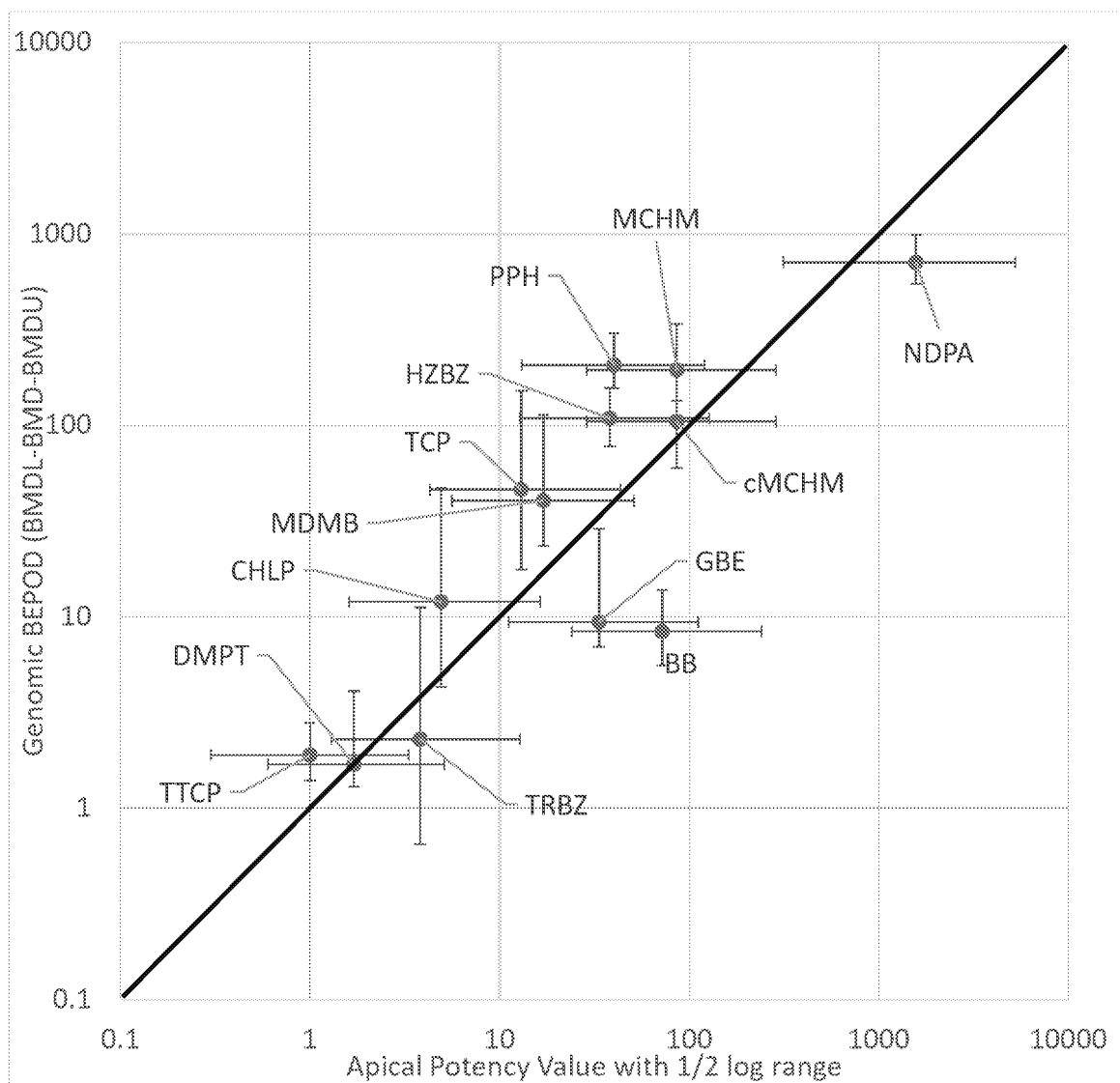
Figure 14. Comparison of the Most Sensitive Apical ½ Log Potency Range to the Most Sensitive GO Biological Processes BEPOD

Data from Figure 1–Figure 13 in this document were compiled to allow a larger scale comparison of apical and gene set-based biological potency estimates. The most sensitive apical potency values (NOAEL or BMD) from guideline toxicity assessments are plotted on the x-axis and the BEPOD range (BMD$_L$-BMD-BMD$_U$) from the GO Biological Processes analysis from 4- or 5-day GDRS studies are plotted on the y-axis. A diagonal 1-to-1 line is drawn as reference to perfect agreement between the potency values. The points to the left of the line demonstrate more sensitive apical endpoints, whereas those to the right exhibited more sensitive BEPODs. Overall, the apical and BEPOD values strongly agree, as indicated by R^2 = 0.89.
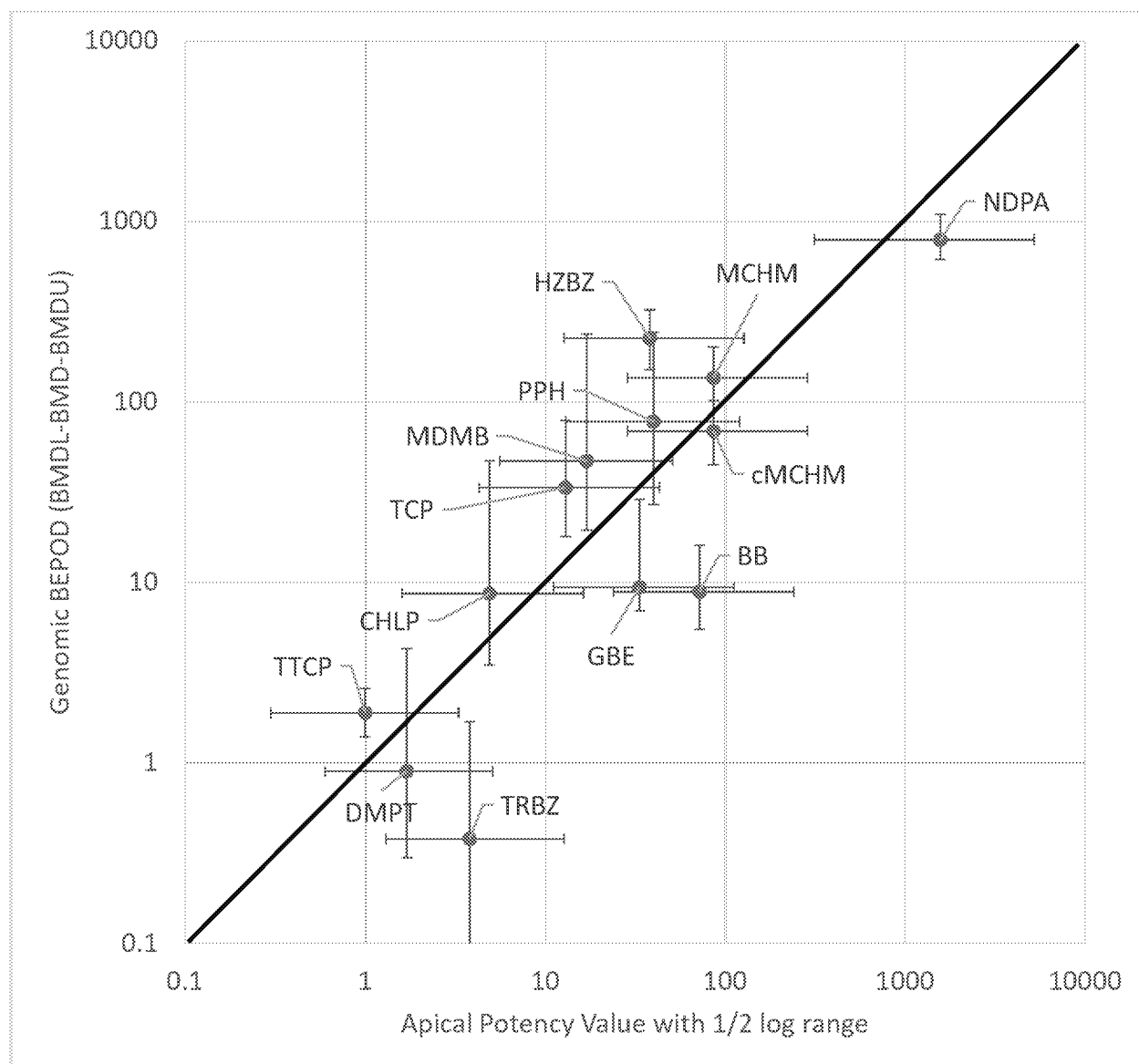
Figure 15. Comparison of the Most Sensitive Apical ½ Log Potency Range to the Most Sensitive MSigDB C2 Gene-Set BEPOD

Data from Figure 1– Figure 13 in this document were compiled to allow a larger scale comparison of apical and gene set-based biological potency estimates. The most sensitive apical potency values (NOAEL or BMD) from guideline toxicity assessments are plotted on the x-axis and the BEPOD range (BMD_L-BMD-BMD_U) from the MSigDB C2 gene-set analysis from 4- or 5-day GDRS studies are plotted on the y-axis. A diagonal 1-to-1 line is drawn as a reference to perfect agreement between the types of potency values. The points to the left of the line demonstrate more sensitive apical endpoints, whereas those to the right exhibited more sensitive BEPODs. Overall, the apical and BEPOD values as indicated by $R^2 = 0.92$ strongly agree.

Appendix A References

1. NTP (National Toxicology Program). 2013. Toxicology and carcinogenesis studies of ginkgo biloba extract (CAS No. 90045-36-6) in F344/N rats and B6C3F1/N mice (gavage studies). https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr578_508.pdf.

2. Gene Ontology Consortium The Gene Ontology project in 2008. Nucleic Acids Res. 2008; 36(Database issue):D440-444. http://dx.doi.org/10.1093/nar/gkm883

3. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011; 27(12):1739-1740. http://dx.doi.org/10.1093/bioinformatics/btr260

4. NTP (National Toxicology Program). 2015. West Virginia chemical spill: Prenatal developmental toxicity study June 2015 NTP update. https://ntp.niehs.nih.gov/ntp/research/areas/wvspill/prenatal_wvupdate_june2015_508.pdf.

5. NTP (National Toxicology Program). 2016. West Virginia chemical spill: 5-day rat toxicogenomic studies July 2016 NTP update. https://ntp.niehs.nih.gov/ntp/research/areas/wvspill/tgmx_update_july2016_508.pdf.

6. Robust summary of a sub-chronic toxicity study of 1-phenoxypropan-2-ol (CASRN: 770-35-4) performed by the oral route. 1997. https://echa.europa.eu/registration-dossier/-/registered-dossier/14251/7/6/2.

7. NTP (National Toxicology Program). 1994. NTP toxicology and carcinogenesis studies of tricresyl phosphate (CAS No. 1330-78-5) in F344/N rats and B6C3F1 mice (gavage and feed studies). https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr433.pdf.

8. NTP (National Toxicology Program). 2012. Toxicology and carcinogenesis studies of N,N-dimethyl-p-toluidine (CAS No. 99-97-8) in F344/N rats and B6C3F1/N mice (gavage studies). https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr579_508.pdf.

9. Dunnick JK, Shockley KR, Morgan DL, Brix A, Travlos GS, Gerrish K, Michael Sanders J, Ton TV, Pandiri AR Hepatic transcriptomic alterations for N,N-dimethyl-p-toluidine (DMPT) and p-toluidine after 5-day exposure in rats. Arch Toxicol. 2017; 91(4):1685-1696. http://dx.doi.org/10.1007/s00204-016-1831-7

10. Dodd DE, Pluta LJ, Sochaski MA, Banas DA, Thomas RS Subchronic hepatotoxicity evaluation of bromobenzene in Fischer 344 rats. J Appl Toxicol. 2013; 33(5):370-377. http://dx.doi.org/10.1002/jat.2732

11. Thomas RS, Wesselkamper SC, Wang NC, Zhao QJ, Petersen DD, Lambert JC, Cote I, Yang L, Healy E, Black MB et al. Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. Toxicol Sci. 2013; 134(1):180-194. http://dx.doi.org/10.1093/toxsci/kft094

12. NTP (National Toxicology Program). 1978. Bioassay of hydrazobenzene for possible carcinogenicity. https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr092.pdf.

13. Dodd DE, Pluta LJ, Sochaski MA, Wall HG, Thomas RS Subchronic hepatotoxicity evaluation of hydrazobenzene in Fischer 344 rats. International Journal of Toxicology. 2012; 31(6):564-571. http://dx.doi.org/10.1177/1091581812465322

14. Dodd DE, Pluta LJ, Sochaski MA, Funk KA, Thomas RS Subchronic hepatotoxicity evaluation of 1,2,4-tribromobenzene in Sprague-Dawley rats. International Journal of Toxicology. 2012; 31(3):250-256. http://dx.doi.org/10.1177/1091581812437974

15. Dodd DE, Pluta LJ, Sochaski MA, Banas DA, Thomas RS Subchronic hepatotoxicity evaluation of 2,3,4,6-tetrachlorophenol in Sprague Dawley rats. Journal of Toxicology. 2012; 2012:1-10. http://dx.doi.org/10.1155/2012/376246

16. Dodd DE, Pluta LJ, Sochaski MA, Funk KA, Thomas RS Subchronic thyroid toxicity evaluation of 4,4'-methylenebis(N, N'-dimethyl)aniline in Fischer 344 rats. Journal of Toxicology and Environmental Health - Part A: Current Issues. 2012; 75(11):637-648. http://dx.doi.org/10.1080/15287394.2012.688481

17. NCI (National Cancer Institute). 1979. Bioassay of 4,4'-methylenebis-(N, N-Dimethyl)venzenamine for possible carcinogenicity.

18. NTP (National Toxicology Program). 1998. NTP toxicology and carcinogenesis studies of chloroprene (CAS No. 126-99-8) in F344/N rats and B6C3F1 mice (inhalation studies). https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr467.pdf.

19. Thomas RS, Himmelstein MW, Clewell HJ, 3rd, Yang Y, Healy E, Black MB, Andersen ME Cross-species transcriptomic analysis of mouse and rat lung exposed to chloroprene. Toxicol Sci. 2013; 131(2):629-640. http://dx.doi.org/10.1093/toxsci/kfs314

20. Dodd DE, Pluta LJ, Sochaski MA, Funk KA, Thomas RS Subchronic urinary bladder toxicity evaluation of N-Nitrosodiphenylamine in Fischer 344 rats. Journal of Applied Toxicology. 2013; 33(5):383-389. http://dx.doi.org/10.1002/jat.2798

21. NCI (National Cancer Institute). 1979. Bioassay of N-Nitrosodiphenylamine for Possible Carcinogenicity.

National Toxicology Program

NTP Central Data Managment, MD K2-05
National Institute of Enviromental Health Sciences
P.O. Box 12233
Research Triangle Park, NC 27709

http://ntp.niehs.nih.gov